

Bounds on the capacity of OFDM underspread frequency selective fading channels

Itsik Bergel^{†*}, Sergio Benedetto[‡]

Abstract

The analysis of the channel capacity in the absence of prior channel knowledge (noncoherent channel) has gained increasing interest in recent years, but it is still unknown for the general case. In this paper we derive bounds on the capacity of the noncoherent, *underspread* complex Gaussian, orthogonal frequency division multiplexing (OFDM), wide sense stationary channel with uncorrelated scattering (WSSUS), under a peak power constraint or a constraint on the second and fourth moments of the transmitted signal. These bounds are characterized only by the system signal-to-noise ratio (SNR) and by a newly defined quantity termed *effective coherence time*. Analysis of the effective coherence time reveals that it can be interpreted as the length of a block in the block fading model in which a system with the same SNR will achieve the same capacity as in the analyzed channel. Unlike commonly used coherence time definitions, it is shown that the effective coherence time depends on the SNR, and is a nonincreasing function of it.

We show that for low SNR the capacity is proportional to the effective coherence time, while for higher SNR the coherent channel capacity can be achieved provided that the effective coherence time is large enough.

I. INTRODUCTION

The analysis of communication systems is often performed under the assumption of perfect channel knowledge. In practical systems, however, the channel needs to be estimated from the communication signal itself, making this assumption unrealistic.

[†] School of Engineering, Bar-Ilan University, 52900 Ramat-Gan, Israel e-mail: bergeli@eng.biu.ac.il

[‡] Dipartimento di Elettronica - Politecnico di Torino e-mail: benedetto@polito.it

This work has been funded by PRIMO, a research project financed by MIUR, the Italian Ministry of Education and Research

* Corresponding author.

In this paper we study the capacity of the noncoherent, underspread, complex Gaussian, orthogonal frequency division multiplexing (OFDM), wide sense stationary channel with uncorrelated scattering (WSSUS), under a peak power constraint or a constraint on the second and fourth moments of the transmitted signal (commonly termed *quadratic power constraint*). We use the term noncoherent channel capacity to describe the capacity of the channel when neither the transmitter nor the receiver have any prior knowledge on the channel realization, but both have exact knowledge on the channel statistics.

The OFDM model provides a simple representation of an underspread frequency selective channel [1]. In underspread channels the channel delay spread is significantly smaller than the channel coherence time [2]. Choosing the OFDM symbol length to be significantly larger than the channel delay spread, and yet significantly smaller than the channel coherence time, results in a useful approximation of the underspread channel which is both accurate and convenient for the analysis. (Detailed description on the connection between the OFDM model and the general WSSUS channel can be found in [3], [4].)

This paper presents novel bounds on the capacity of the noncoherent underspread WSSUS channel that are characterized only by the system signal-to-noise ratio (SNR) and by a newly defined quantity termed *effective coherence time*. We show that the bounds are good for almost all of the SNR range as long as the effective coherence time is significantly larger than the OFDM symbol length. Analysis of the effective coherence time reveals that it characterizes the system ability to estimate the channel, and can be interpreted as the length of a block in the block fading model that achieves the same capacity as the analyzed channel at the same SNR. Surprisingly, and unlike standard coherence time definitions, the effective coherence time is actually a function of the system SNR. This dependence of the effective coherence time on SNR stems from the fact that at high SNR the system is more sensitive to changes in the channel, and hence will effectively see a shorter coherence time.¹ The paper also includes a detailed study of the relation between the effective coherence time and the system SNR.

The noncoherent underspread WSSUS channel capacity was well characterized for the high

¹ A simple intuition for the decrease of coherence time with SNR is as follows. Consider an intuitive definition of coherence time for stationary channels: The coherence time is the maximal time interval for which the channel auto-correlation function does not decrease below a pre-specified threshold. The question is how to select the threshold value. Intuitively, a more sensitive system (higher SNR) will choose a higher threshold value, and hence will see a shorter coherence time.

SNR limit and for the low SNR limit (which is equivalent to the large bandwidth limit). For the low SNR limit, the capacity of the noncoherent underspread channel was shown to be proportional to $p_x^2 T_c / 2$, where p_x is the SNR and T_c is the channel coherence time [3]–[7]. These results were obtained for different channel models with different constraints. Medard and Gallager [5] analyzed a stationary wideband channel with quadratic power constraint, and presented a novel definition of the channel coherence time which is calculated from the channel correlation function. In this case, the proportionality constant is the ratio between the transmitted signal fourth moment and the square of its second moment (also termed *Kurtosis*). The results in that work were derived for the wide bandwidth limit (in which the capacity depends on the number of resolvable channel taps, see also [8], [9]). In this paper, the channel bandwidth is finite, and we address the above result only in its low SNR limit interpretation.

Sethuraman and Hajek [6] considered the block fading model with a peak power constraint. In that case the channel coherence time is the length of a block, and the proportionality constant was shown to be the ratio between the peak and average powers. They also considered a flat fading stationary channel model (i.e., without multipath spread) and achieved equivalent results, but using measures of the channel spectral distribution function (with no definition of channel coherence time). In a later work together with Wang and Lapidot [7], they considered delay-separable frequency selective fading channels, and showed that the capacity is actually proportional to $p_x^2 (T_c - 1) / 2$; here the definition of the channel coherence time is the identical to the one used by Medard and Gallager [5] but the term coherence time is not used. (Note that for underspread channels the difference between T_c and $T_c - 1$ is negligible.)

Durisi et al. used a similar channel model, but paid more attention to the discretization of the continuous time channel (and did not require the channel to be delay separable). As a result, they had worked with a time-frequency channel transfer function (which resembles in nature to the OFDM model used herein). They also did not use the term coherence time, but their results revealed the same behavior of the capacity in the low SNR limit, using a channel measure which is an extension of Medard and Gallager [5] coherence time to the time-frequency channel. (In their work they also analyzed the capacity with per frequency peak power constraint. This power constraint is not discussed herein, as it is quite different and leads to a significantly different capacity behavior.)

For the high SNR limit, the main issue is whether the channel estimation error can decrease

to 0 as the SNR grows to infinity. The channel fading is termed *regular* if it cannot be completely predicted based on the full knowledge of its past [10]. An alternative definition uses the covariance matrix of the channel fading, and the channel is said to be regular if this matrix is full rank [11]. For regular channels, at high enough SNR, the capacity in the absence of channel knowledge is significantly lower than the capacity with perfect channel knowledge, and grows only double logarithmically with SNR [12], [13], [14]. On the other hand, for irregular channels the capacity can continue to grow logarithmically with SNR, (with possible degradation in the pre log constant) [13], [14].

Interestingly, no capacity expression directly depends on the common definition of coherence time (e.g., [15]), i.e., the inverse of the channel Doppler spread. For stationary channels, the only coherence time definition that was shown to be proportional to the capacity is the one given by Medard and Gallager [5], which only applies in the low SNR limit. The definition of effective coherence time in this paper coincides with Medard's and Gallager's coherence time definition in the low SNR limit, and extends it to higher SNR. Thus, the effective coherence time can characterize the channel capacity for all SNRs.

The paper is organized as follows: the notation and channel model are presented in Sections II and III respectively. The main results, including the definition of the effective coherence time and the capacity bounds are given in Section IV. A discussion of the results including the properties of the effective coherence time, analysis of the bounding gap, numerical examples and comparison to known results are given in Section V. Note that the proof of each of the theorems (providing capacity bounds) is partly based on information theory and partly on estimation theory. For convenience, we collect those parts in different sections. Hence, Theorem n ($n = 1, 2, 3$) is proved by lemma $n.a$ in Section VI (information theory part) and lemma $n.b$ in Section VII (estimation theory part). Concluding remarks are given in Section VIII.

II. NOTATION

Throughout the paper we use Roman boldface lower case and upper case letters to denote random scalars and vectors, respectively (e.g., \mathbf{x} , \mathbf{X}). Roman italic lower case and upper case letters are used for deterministic scalars and vectors, respectively (e.g., x , X). Deterministic matrices are represented by sans-serif letters and random matrices by calligraphic letters (e.g., \mathbf{X} , \mathcal{X}). Exceptions to this rule are scalar quantities that are commonly denoted by uppercase

letters. These include the channel parameters N and L , the coherence time symbol (T_c , $\hat{T}_c(\cdot)$ and \hat{T}_{c_0}), and the capacity symbol C . Another exception are operators that can result in a deterministic or random quantity according to their operand, i.e., the spectrum operator \mathcal{S} . (see definition below) and the symbol set operator $\mathcal{I}(\cdot)$ (see definition in Lemma 1.b).

The identity matrix is denoted by \mathbf{I} , 0_N and 1_N denote the $N \times 1$ vectors of all zeros and all ones, respectively. $\mathbf{A} \otimes$ symbolizes the Kronecker product, a \dagger stands for transposition and complex conjugation, $\|\mathbf{A}\|$ denotes the absolute value of the determinant of the matrix \mathbf{A} , and $\text{diag}(X)$ is the diagonal matrix with the elements of the vector X on its diagonal. The vector stacking is defined as: $X_k^n \triangleq [X_k^T, \dots, X_n^T]^T$. The b, d element of a matrix \mathbf{A} is denoted as $(\mathbf{A})_{b,d}$ and the b -th element of the vector X_k is denoted as $x_{k,b}$.

For discrete Fourier transform (DFT) analysis, \mathbf{F} denotes the DFT matrix with elements:

$$F_{m,n} = \frac{1}{\sqrt{N}} e^{-2j\pi mn/N}, \quad (1)$$

$\tilde{X} = \mathbf{F}^\dagger X$ denotes the DFT of a vector, and each element in this vector is termed a frequency bin. The spectrum of a vector is marked by:

$$\mathcal{S}_X = \text{diag}(\tilde{X}) \text{diag}(\tilde{X})^\dagger. \quad (2)$$

The cross-covariance matrix of the vectors \mathbf{X} and \mathbf{Y} is denoted as:

$$\mathbf{C}_{\mathbf{X},\mathbf{Y}} = E \left[(\mathbf{X} - E[\mathbf{X}]) (\mathbf{Y} - E[\mathbf{Y}])^\dagger \right], \quad (3)$$

and the auto-covariance matrix of the vector \mathbf{X} is denoted as $\mathbf{C}_\mathbf{X} = \mathbf{C}_{\mathbf{X},\mathbf{X}}$.

Finally, any matrix ordering in this paper is in the positive definite sense, i.e., $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semi-definite.

III. SYSTEM MODEL

A. Channel model

We use the OFDM model [1], [4], [16], [17], in which the DFT of the received signal is given by:

$$\begin{aligned} \tilde{\mathbf{Y}}_k &= \sqrt{N} \text{diag}(\tilde{\mathbf{H}}_k) \tilde{\mathbf{X}}_k + \tilde{\mathbf{W}}_k \\ &= \sqrt{N} \text{diag}(\tilde{\mathbf{X}}_k) \tilde{\mathbf{H}}_k + \tilde{\mathbf{W}}_k \end{aligned} \quad (4)$$

where $\tilde{\mathbf{X}}_k \triangleq \mathbf{F}^\dagger \mathbf{X}_k$ and $\tilde{\mathbf{H}}_k \triangleq \mathbf{F}^\dagger \mathbf{H}_k$ are the DFT of the input vector and channel impulse response at the k -th symbol respectively, and both are vectors of length N . Each input vector constitutes an OFDM symbol. The frequency domain noise $\tilde{\mathbf{W}}_k \triangleq \mathbf{F}^\dagger \mathbf{W}_k$ is a zero mean complex Gaussian vector with covariance matrix: $\mathbf{C}_{\tilde{\mathbf{W}}} = \mathbf{I}$. The multiplication by \sqrt{N} follows from the definition of the DFT matrix, (1). The channel itself is better characterized in the time domain where $\mathbf{H}_k = [\mathbf{h}_{k,0}, \dots, \mathbf{h}_{k,L-1}, 0, \dots, 0]^T$, and L is termed the channel delay spread.

This model is an appropriate representation of an underspread channel in which we assume $T_c \gg N \gg L - 1$,² where T_c is the channel coherence time [15]. The assumption guarantees that the channel Doppler spread is small enough to be negligible, and that the channel impulse response can be considered constant throughout the N samples that constitute an OFDM symbol (for analysis of faster fading channels see [18]). The small channel delay spread, L , permits to neglect the time required for cyclic prefix ($L - 1$ samples after each OFDM symbol).

In Section IV we define a new parameter, the *effective coherence time*, $\hat{T}_c(p_x)$, which identifies the effective number of received samples that can be used for channel estimation. The impact of the effective coherence time on the channel capacity will be treated in later sections.

The mathematical analysis in the paper requires the following assumptions:

Assumption I: The channel is underspread, i.e., the channel coherence time and the effective coherence time (measured in channel samples) are large enough so that: $T_c, \hat{T}_c(p_x) \gg N$.

Assumption II: The channel has a proper complex Gaussian distribution ([19]).

Assumption III: The transmitter has no knowledge on the channel realization.

Assumption IV: The channel is ergodic, (wide sense) stationary, with uncorrelated scattering (WSSUS channel).

Assumption V: The auto-correlation function of the different channel taps is identical up to a multiplication by a scalar.

Assumption VI: The expectation of the channel impulse response is frequency flat.

Assumption I is actually not needed for the analysis presented herein. We state it as our first assumption because it is the basis for the OFDM model. If the coherence time is not significantly larger than the symbol length, then our model, which implicitly assumes that the channel does not change during an OFDM symbol, will not be realistic.

²We use the notation $N \gg L - 1$ in order to emphasize that in the flat fading case ($L = 1$) it is sufficient to use $N = 1$

Assumption III stems from the fact that the system has no feedback. As a result, the transmitted messages are statistically independent from the channel. The requirement of uncorrelated scattering in Assumption IV is needed mostly for low SNR where we show that it is better to concentrate the transmission power in a single frequency bin. The uncorrelated scattering guarantees that if in one OFDM symbol we transmit in a certain frequency bin, the best channel estimate for the next OFDM symbol will be achieved at the same frequency bin (see derivation in Appendix B).

Assumptions V and VI somewhat narrow the scope of the analysis. However, we keep them to simplify the mathematical derivation, and to reach closed-form expressions. Assumption V, although not necessarily realistic, is common to many simple channel models (e.g., assuming that all multipath components have identical doppler spread). Channels that satisfy this assumption are sometimes referred to as *delay separable channels* [7]. Assumption VI basically states that at most one channel tap has an expectation different from zero, and it includes the case of Rayleigh fading (the expectation is zero) and the case of line of sight (LOS) propagation (one channel tap has an expectation different from zero).

Using Assumption II, the stacking of the channel vector has a proper complex Gaussian distribution, $\mathbf{H}_0^k \sim \mathcal{CN}(E[\mathbf{H}_0^k], \mathbf{C}_{\mathbf{H}_0^k})$. Using Assumption IV the amplitudes of the different channel taps (in the time domain, $\mathbf{h}_{k,0}, \dots, \mathbf{h}_{k,L-1}$) are statistically independent. Using also Assumption V the covariance matrices of the channel taps are identical up to a multiplication by a scalar. Considering $k+1$ OFDM symbols, we denote by Δ_k the single-tap $(k+1) \times (k+1)$ Toeplitz covariance matrix, in which each element is given by

$$(\Delta_k)_{i,j} = \mathbf{C}_{\mathbf{h}_{i,0}, \mathbf{h}_{j,0}} / \mathbf{C}_{\mathbf{h}_{0,0}}. \quad (5)$$

We also define the $k \times 1$ vector D_k in which the i -th element is given by

$$d_{k,i} = (\Delta_k)_{i,k}, \quad i = 0, \dots, k-1. \quad (6)$$

This vector is the correlation between the channel tap value at the k -th symbol and its value at all previous symbols. Considering all channel taps, the covariance matrix of the stacked channel

vector \mathbf{H}_0^k is:

$$\mathbf{C}_{\mathbf{H}_0^k} = \Delta_k \otimes \mathbf{C}_{\mathbf{H}_0} = \begin{pmatrix} (\Delta_k)_{0,0}\mathbf{C}_{\mathbf{H}_0} & (\Delta_k)_{0,1}\mathbf{C}_{\mathbf{H}_0} & \dots & (\Delta_k)_{0,k}\mathbf{C}_{\mathbf{H}_0} \\ (\Delta_k)_{1,0}\mathbf{C}_{\mathbf{H}_0} & (\Delta_k)_{1,1}\mathbf{C}_{\mathbf{H}_0} & \dots & (\Delta_k)_{1,k}\mathbf{C}_{\mathbf{H}_0} \\ \vdots & \vdots & \ddots & \vdots \\ (\Delta_k)_{k,0}\mathbf{C}_{\mathbf{H}_0} & (\Delta_k)_{k,1}\mathbf{C}_{\mathbf{H}_0} & \dots & (\Delta_k)_{k,k}\mathbf{C}_{\mathbf{H}_0} \end{pmatrix}, \quad (7)$$

and $\mathbf{C}_{\mathbf{H}_0}$ is an $N \times N$ diagonal matrix, in which each diagonal element represents the power of one channel tap (note that due to Assumption IV, we can write $\mathbf{C}_{\mathbf{H}_k} = \mathbf{C}_{\mathbf{H}_0}$ for any k). In the following we mostly consider the frequency domain channel, $\tilde{\mathbf{H}}_k = \mathbf{F}^\dagger \mathbf{H}_k$. Its covariance matrix is given by $\mathbf{C}_{\tilde{\mathbf{H}}_k} = \mathbf{F}^\dagger \mathbf{C}_{\mathbf{H}_k} \mathbf{F}$. For normalization purposes we will use $\text{Tr}[\mathbf{C}_{\mathbf{H}_0}] = 1$ (and hence all elements on the diagonal of $\mathbf{C}_{\tilde{\mathbf{H}}_0}$ equal $1/N$).

Another quantity which is useful in the characterization of the channel is the channel conditional covariance matrix given the past transmitted and received symbols:

$$\tilde{\mathcal{E}}_k = E \left[(\tilde{\mathbf{H}}_k - E[\tilde{\mathbf{H}}_k])(\tilde{\mathbf{H}}_k - E[\tilde{\mathbf{H}}_k])^\dagger | \tilde{\mathbf{X}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1} \right]. \quad (8)$$

This conditional covariance matrix is in general a random quantity since it depends on the random vectors $\tilde{\mathbf{X}}_0^{k-1}$ (as the channel and the noise are jointly Gaussian, it does not depend on $\tilde{\mathbf{Y}}_0^{k-1}$). In some cases (e.g., constant amplitude modulations) the resulting conditional covariance matrix is deterministic (does depend on $\tilde{\mathbf{X}}_0^{k-1}$), and hence will be denoted by $\tilde{\mathcal{E}}_k$. We will also use the limit as both time and SNR tend to infinity $\mathcal{E}_\infty = \lim_{k \rightarrow \infty} \lim_{p_x \rightarrow \infty} \tilde{\mathcal{E}}_k$. Further details on the conditional channel distribution given the past transmitted and received symbols can be found in Section VII.

B. Power constraints

The input signal has constraints both on its average power and on its “peakiness”. Defining the OFDM symbol power and power matrix as:

$$\mathbf{p}_k = \tilde{\mathbf{X}}_k^\dagger \tilde{\mathbf{X}}_k, \quad \mathcal{P}_k = \text{diag}([\mathbf{p}_0, \dots, \mathbf{p}_k]), \quad (9)$$

respectively, we consider two types of constraints preventing the use of very high peak powers. The first, *peak constraint*, limits the peak power of each OFDM symbol:

$$\mathbf{p}_k \leq N p_x, \quad \forall k, \quad (10)$$

with probability 1 (The peak power constraint does not involve statistical averaging, and therefore it is more relevant in practical systems).

The second constraint, *quadratic constraint* [5], [8], limits the first and second moments of the OFDM symbol power:

$$E[\mathbf{p}_k] \leq Np_x, \quad E[\mathbf{p}_k^2] \leq \alpha N^2 p_x^2, \quad \forall k, \quad (11)$$

where $\alpha \geq 1$ is a positive constant. This constraint limits the probability of very high peak powers, but still permits a relatively simple analysis (for example by allowing Gaussian signaling).

As the noise variance is normalized to 1, in the following we will mostly refer to p_x as the system SNR.

IV. MAIN RESULTS

In this section we introduce without proof upper and lower bounds on the capacity of the noncoherent channel described in the previous section. These bounds are characterized only by the system SNR and the *effective coherence time*:

Definition 1: The *effective coherence time* is given by:

$$\hat{T}_c(p_x) = 2N \lim_{k \rightarrow \infty} D_k^\dagger \left(Np_x \left[\Delta_{k-1} - D_k D_k^\dagger \right] + \mathbf{I} \right)^{-1} D_k + N. \quad (12)$$

As it will be shown, the effective coherence time characterizes the "effective number of channel samples usable for channel estimation". Although the previous statement is not precise now, it will be discussed further in Subsection V-A.

Theorem 1: The capacity of a channel with a peak power constraint is upper bounded by:

$$C \leq \min \left(\text{UB}_{\text{low}}^{(\text{pk})}(p_x), \text{UB}_{\text{coh}}(p_x) \right), \quad (13)$$

and the capacity of a channel with a quadratic power constraint is upper bounded by:

$$C \leq \min \left(\text{UB}_{\text{low}}^{(\text{qd})}(p_x), \text{UB}_{\text{coh}}(p_x) \right) \quad (14)$$

where:

$$\text{UB}_{\text{coh}}(p_x) = E \left[\log \left(1 + Np_x |\tilde{\mathbf{h}}_{0,0}|^2 \right) \right] \quad (15)$$

$$\text{UB}_{\text{low}}^{(\text{pk})}(p_x) = Np_x \left| E \left[\tilde{\mathbf{h}}_{0,0} \right] \right|^2 + p_x - \frac{1}{N} \log \left(1 + \frac{Np_x}{1 + \frac{p_x}{2} (\hat{T}_c(p_x) - N)} \right) \quad (16)$$

$$\text{UB}_{\text{low}}^{(\text{qd})}(p_x) = Np_x \left| E \left[\tilde{\mathbf{h}}_{0,0} \right] \right|^2 + \alpha \frac{p_x^2 (\hat{T}_c(p_x) - N)}{2 + p_x (\hat{T}_c(p_x) - N)} + \frac{1}{2} \alpha N p_x^2. \quad (17)$$

Proof of Theorem 1: see Result 1, Lemma 1.a and Lemma 1.b.³

The term $\text{UB}_{\text{coh}}(p_x)$ in the bound is the capacity of the coherent channel with only average power constraint. We will show in Subsection V-B that it is useful in the medium-to-high SNR regime. For low SNR the coherent channel capacity is not achievable and tighter bounds are described by $\text{UB}_{\text{low}}^{(\text{pk})}(p_x)$ and $\text{UB}_{\text{low}}^{(\text{qd})}(p_x)$. For low enough SNR, using $\log(1+x) = x + \frac{1}{2}x^2 + o(x^2)$ and $1/(1+x) = 1 - x + o(x)$, these bounds can be approximated as:

$$\text{UB}_{\text{low}}^{(\text{pk})}(p_x) \approx Np_x \left| E \left[\tilde{\mathbf{h}}_{0,0} \right] \right|^2 + \frac{p_x^2}{2} \hat{T}_c(p_x), \quad (18)$$

$$\text{UB}_{\text{low}}^{(\text{qd})}(p_x) \approx Np_x \left| E \left[\tilde{\mathbf{h}}_{0,0} \right] \right|^2 + \alpha \frac{p_x^2}{2} \hat{T}_c(p_x). \quad (19)$$

If the channel has a constant term ($|E[\tilde{\mathbf{h}}_{0,0}]|$) different from zero, this is the dominant term. Otherwise, the bounds are proportional to $p_x^2 \hat{T}_c(p_x)$, but the quadratic power constraint allows a capacity which is α times larger.

For the lower bounds, we choose specific input distributions (modulations). For low SNR it is important to use a modulation that maximizes the estimation accuracy subject to the power constraints. Such a maximization is achieved by constant amplitude modulation, and therefore we derive a lower bound using Quadrature Phase Shift Keying (QPSK). As it was noted in previous works (e.g., [6], [9]), in the low SNR regime it is better to use only part of the time and/or part of the frequency band. Using only part of the available degrees of freedom reduces the number of parameters that need to be estimated and hence results in better estimation. In the following bounds we allow the use of only r out of the N frequency bins, and allow transmitting in $1/\beta$ of the time. In the time domain the signal is transmitted in long blocks (significantly larger than the effective coherence time), and the silent periods between blocks results in a duty cycle of $1/\beta$. The resulting bound is:

Theorem 2: The capacity of a channel with a peak power constraint is lower bounded by:

$$C \geq \max_{r \leq N} \text{LB}_{\text{QPSK}}(p_x, r, 1) \quad (20)$$

³The proof of each theorem is divided into two lemmas. In the first lemma we use information-theoretical arguments to prove the theorem assuming a knowledge of the conditional channel distribution (defined by the conditional channel covariance matrix), while in the second lemma we bound the conditional channel covariance matrix.

and the capacity of a channel with a quadratic power constraint is lower bounded by:

$$C \geq \max_{r \leq N} \max_{1 \leq \beta \leq \alpha} \text{LB}_{\text{QPSK}}(p_x, r, \beta) \quad (21)$$

where

$$\text{LB}_{\text{QPSK}}(p_x, r, \beta) = \begin{cases} \frac{r}{N\beta} E \left[C_{\text{QPSK}} \left(\frac{\left| \sqrt{NE} [\tilde{\mathbf{h}}_{0,0}] + \sqrt{1 - \frac{1}{1 + \frac{\beta}{r} \frac{p_x}{2} (\hat{T}_c(\frac{\beta}{r} p_x) - N)}} \mathbf{u} \right|^2}{\frac{r}{N} \frac{1}{\beta p_x} + \frac{1}{1 + \frac{\beta}{r} \frac{p_x}{2} (\hat{T}_c(\frac{\beta}{r} p_x) - N)}}} \right) \right] & r < N \\ \frac{1}{\beta} E \left[C_{\text{QPSK}} \left(\frac{\left| \sqrt{NE} [\tilde{\mathbf{h}}_{0,0}] + \sqrt{1 - \frac{1}{1 + \frac{\beta}{L} \frac{p_x}{2} (\hat{T}_c(\beta p_x) - N)}} \mathbf{u} \right|^2}{\frac{1}{\beta p_x} + \frac{1}{1 + \frac{\beta}{L} \frac{p_x}{2} (\hat{T}_c(\beta p_x) - N)}}} \right) \right] & r = N \end{cases}, \quad (22)$$

$$C_{\text{QPSK}}(\rho) = \log 4 - \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \log(1 + e^{-2(\rho + \sqrt{\rho} y)}) dy \quad (23)$$

is the achievable mutual information for QPSK transmission over AWGN channel with SNR equal to ρ ([20]) and $\mathbf{u} \sim \mathcal{CN}(0, 1)$ is a proper complex Gaussian random variable.

Proof of Theorem 2: see Lemma 2.a and Lemma 2.b.

Note that if the channel has zero mean ($E[\tilde{\mathbf{H}}_0] = 0$) then the bound in (22) simplifies to:

$$\text{LB}_{\text{QPSK}}(p_x, r, \beta) = \begin{cases} \frac{r}{N\beta} E \left[C_{\text{QPSK}} \left(\frac{\frac{N\beta^2}{r^2} \frac{p_x^2}{2} (\hat{T}_c(\frac{\beta}{r} p_x) - N) |\mathbf{u}|^2}{1 + \frac{N}{r} \beta p_x + \frac{\beta}{r} \frac{p_x}{2} (\hat{T}_c(\frac{\beta}{r} p_x) - N)} \right) \right] & r < N \\ \frac{1}{\beta} E \left[C_{\text{QPSK}} \left(\frac{\frac{\beta^2}{L} \frac{p_x^2}{2} (\hat{T}_c(\beta p_x) - N) |\mathbf{u}|^2}{1 + \beta p_x + \frac{\beta}{L} \frac{p_x}{2} (\hat{T}_c(\beta p_x) - N)} \right) \right] & r = N \end{cases}.$$

This bound is especially interesting in the low SNR limit. In that case, the best bound uses only a single frequency bin, and the minimum allowed duty cycle. Using $C_{\text{QPSK}}(x) = x + o(x)$ and $1/(1+x) = 1 + o(1)$, the low SNR approximation of the above bound for the peak power constraint is:

$$\text{LB}_{\text{QPSK}}(p_x, 1, 1) \approx N p_x \left| E[\tilde{\mathbf{h}}_{0,0}] \right|^2 + \frac{p_x^2}{2} (\hat{T}_c(p_x) - N), \quad (24)$$

and the low SNR approximation for the quadratic power constraint is:

$$\text{LB}_{\text{QPSK}}(p_x, 1, \alpha) \approx N p_x \left| E[\tilde{\mathbf{h}}_{0,0}] \right|^2 + \alpha \frac{p_x^2}{2} (\hat{T}_c(\alpha p_x) - N). \quad (25)$$

Comparing with (18) and (19) we see that the bounds are tight for the peak power constraint as long as $\hat{T}_c(p_x) \gg N$,⁴ while for the quadratic power constraint the bound tightness depends

⁴Much of the work on OFDM is concerned with the peak-to-average power ratio (PAPR), while in this work we only consider the entire symbol energy. Yet, we note that in the low SNR limit the lower bound uses a single frequency bin, and hence the PAPR is 1.

on the properties of the effective coherence time. When the limit $\hat{T}_{c_0} = \lim_{p_x \rightarrow 0} \hat{T}_c(p_x)$ exists, the bounds are tight up to a factor of $(\hat{T}_{c_0} - N)/\hat{T}_{c_0}$. We will further discuss properties of the effective coherence time in Subsection V-A.

For higher SNR we expect the coherent channel bound to be achievable. In the coherent channel the capacity is achieved using a wideband Gaussian modulation, and we expect it to be close to optimal also for the noncoherent channel, in the appropriate SNR regime. Several simulations performed using Gaussian modulation showed convergence to the upper bound, and in Subsection V-A we even discuss a simulation-based approximation that holds for large enough effective coherence times and can be used to approximately lower bound the channel capacity using Gaussian modulation.

However, so far we have not been able to prove a useful lower bound using Gaussian modulation. Instead, we present here a bound which is based on truncated complex Gaussian modulation. This modulation is defined in Section VI. In essence, it uses a proper complex Gaussian distribution with the condition of a minimal and maximal power for the signal transmitted in each active bin. The presented bound will show that indeed the coherent channel capacity is achievable for high SNR and large enough effective coherence times.

For this bound we also require a zero mean channel ($E[\tilde{\mathbf{h}}_{0,m}] = 0$). An alternative (simpler) bound that does not require this assumption is presented in Appendix D. This alternative bound is in general less tight, and therefore we prefer to assume a zero mean channel and focus on the following bound:

Theorem 3: The capacity of a channel with zero mean ($E[\tilde{\mathbf{h}}_{0,m}] = 0$) and a quadratic power constraint is lower bounded by:

$$C \geq \sup_{\eta > 0} \max_{r \leq N} \max_{1 \leq \beta \leq \frac{\alpha r(1+\eta)^2}{1+r(1+\eta)^2}} \text{LB}_{\text{TG}}^{(\text{qd})}(p_x, r, \beta, \eta) - \frac{r}{N\beta} \eta \quad (26)$$

and the capacity of a channel with zero mean ($E[\tilde{\mathbf{h}}_{0,m}] = 0$) and a peak power constraint is lower bounded by:

$$C \geq \sup_{\eta > 0} \sup_{\xi > \eta} \max_{r \leq N} \text{LB}_{\text{TG}}^{(\text{pk})}(p_x, r, \eta, \xi) + \frac{r}{N} \log(e^{-\eta} - e^{-\xi}) \quad (27)$$

where

$$\text{LB}_{\text{TG}}^{(\text{qd})}(p_x, r, \beta, \eta) = \begin{cases} \frac{r}{N\beta} E \left[\log \left(1 + \frac{\frac{N}{r} \beta p_x |\mathbf{u}|^2}{1 + \frac{\frac{N}{r} \beta p_x}{\frac{\beta p_x}{\nu_{\text{qd}}(\eta)} \frac{1}{2r} \left(\hat{T}_c \left(\frac{\beta p_x}{\nu_{\text{qd}}(\eta)} \right) - N \right)}} \right) \right] & r < N \\ \frac{1}{\beta} E \left[\log \left(1 + \frac{\beta p_x |\mathbf{u}|^2}{1 + \frac{\beta p_x}{\frac{\beta p_x}{\nu_{\text{qd}}(\eta)} \frac{1}{2L} \left(\hat{T}_c \left(\frac{\beta p_x}{\nu_{\text{qd}}(\eta)} \right) - N \right)}} \right) \right] & r = N \end{cases}, \quad (28)$$

$$\text{LB}_{\text{TG}}^{(\text{pk})}(p_x, r, \eta, \xi) = \begin{cases} \frac{r}{N} E \left[\log \left(1 + \frac{\frac{N}{r} \frac{p_x}{\xi} \left(1 + \frac{\eta e^{-\eta} - \xi e^{-\xi}}{e^{-\eta} - e^{-\xi}} \right) |\mathbf{u}|^2}{1 + \frac{\frac{N}{r} \frac{p_x}{\xi} \left(1 + \frac{\eta e^{-\eta} - \xi e^{-\xi}}{e^{-\eta} - e^{-\xi}} \right)}{\frac{p_x}{\nu_{\text{pk}}(\eta, \xi)} \frac{1}{2r} \left(\hat{T}_c \left(\frac{p_x}{\nu_{\text{pk}}(\eta, \xi)} \right) - N \right)}} \right) \right] & r < N \\ E \left[\log \left(1 + \frac{\frac{p_x}{\xi} \left(1 + \frac{\eta e^{-\eta} - \xi e^{-\xi}}{e^{-\eta} - e^{-\xi}} \right) |\mathbf{u}|^2}{1 + \frac{\frac{p_x}{\xi} \left(1 + \frac{\eta e^{-\eta} - \xi e^{-\xi}}{e^{-\eta} - e^{-\xi}} \right)}{\frac{p_x}{\nu_{\text{pk}}(\eta, \xi)} \frac{1}{2L} \left(\hat{T}_c \left(\frac{p_x}{\nu_{\text{pk}}(\eta, \xi)} \right) - N \right)}} \right) \right] & r = N \end{cases}, \quad (29)$$

$$\nu_{\text{qd}}(\eta) = (1 + \eta) \log \left(1 + \frac{1}{\eta} \right) \quad (30)$$

$$\nu_{\text{pk}}(\eta, \xi) = \frac{\xi \log(1 + \frac{1}{\eta})}{1 - e^{\eta - \xi}} \quad (31)$$

and $\mathbf{u} \sim \mathcal{CN}(0, 1)$ is a proper complex Gaussian random variable.

Proof of Theorem 3: see Lemma 3.a and Lemma 3.b.

The bound for the quadratic power constraint is tight for high SNR and high effective coherence time. Together with Theorem 2 it yields a pair of bounds encompassing the range of SNR for which the effective coherence time is large enough. Considering in particular the case of $\beta = 1$ and $r = N$, we can see two penalty terms. A capacity penalty term of η (the second term on the right hand side of (26)), and an estimation penalty term in the denominator of (28). To show the bound tightness we can choose for example

$$\eta = \frac{1}{e^{\sqrt{\hat{T}_c(p_x)}} - 1}. \quad (32)$$

Note that $\hat{T}_c(p_x) \geq N \geq 1$ and hence we have $\eta \leq 1$ and $\nu_{\text{qd}}(\eta) \leq 2 \log \left(1 + \frac{1}{\eta} \right) = 2\sqrt{\hat{T}_c(p_x)}$.

Therefore, the estimation penalty term in (28) satisfies:

$$\frac{1 + \beta p_x}{\frac{\beta p_x}{\nu_{\text{qd}}(\eta)} \frac{1}{2L} \left(\hat{T}_c \left(\frac{\beta p_x}{\nu_{\text{qd}}(\eta)} \right) - N \right)} \leq \frac{1 + p_x}{\frac{p_x}{\nu_{\text{qd}}(\eta)} \frac{1}{2L} \hat{T}_c(p_x)} \leq \frac{1 + p_x}{\frac{p_x}{4L} \sqrt{\hat{T}_c(p_x)}},$$

and becomes negligible for large enough $\hat{T}_c(p_x)$. On the other hand, the capacity penalty term is proportional to η , (32), and is also negligible for large $\hat{T}_c(p_x)$. Thus, for high SNR and large enough effective coherence time, the bound can be approximated as:

$$\sup_{0 < \eta < 1} \text{LB}_{\text{TG}}(p_x, N, 1, \eta) \approx E \left[\log \left(1 + p_x |\mathbf{u}|^2 \right) \right]. \quad (33)$$

Since in this case $\tilde{\mathbf{h}}_{0,0} \sim \mathcal{CN}(0, \frac{1}{N})$, this approximation converges to the coherent channel upper bound $\text{UB}_{\text{coh}}(p_x)$ (15).

For the peak power constraint the situation is more complicated due to the power decrease that is required to allow for the truncated Gaussian modulation to satisfy the peak power constraint. This power penalty term is given by $\frac{1}{\xi} \left(1 + \frac{\eta e^{-\eta} - \xi e^{-\xi}}{e^{-\eta} - e^{-\xi}} \right)$ (in the nominator of (29)). For high enough effective coherence time (using for example η as in (32)) the estimation will be good enough so that η will be negligible and also the denominator of (29) will be very close to 1. In such case, at high SNR and $r = N$ the capacity loss compared to the coherent channel capacity converges to:

$$\log \left(\frac{1}{\xi} \left(1 - \frac{\xi e^{-\xi}}{1 - e^{-\xi}} \right) \right) + \log (1 - e^{-\xi}),$$

which is maximized by $\xi = 1.79$ at a loss of -1.21 nats. Thus, the bound will be tight only for very high SNR, where such capacity loss is negligible, if the effective coherence time $\hat{T}_c(p_x)$ is still large enough.

V. DISCUSSION

A. Effective coherence time

The three bounds in the previous section show that the effective coherence time can very well characterize the noncoherent channel capacity. Yet, its definition:

$$\hat{T}_c(p_x) = 2N \lim_{k \rightarrow \infty} D_k^\dagger \left(N p_x \left[\Delta_{k-1} - D_k D_k^\dagger \right] + \mathbf{I} \right)^{-1} D_k + N.$$

is not intuitive and needs a discussion on its meaning and properties. In the following we present some interesting properties of the effective coherence time, followed by a discussion on the reasoning and motivation for each property. In particular, we will also explain why to use the term *effective coherence time* for this quantity.

Properties of the effective coherence time

- 1) Using wideband constant amplitude modulation (i.e., all OFDM frequency bins are active and all have identical amplitude, $r = N$), the conditional channel covariance matrix, given all past transmitted and received symbols, is a diagonal matrix. The diagonal element that corresponds to any channel tap such that $(\mathbf{C}_{\mathbf{H}_0})_{l,l} > 0$ is given by:

$$\lim_{k \rightarrow \infty} (\mathcal{E}_k)_{l,l}^{-1} = (\mathbf{C}_{\mathbf{H}_0})_{l,l}^{-1} + \frac{p_x}{2} \left(\hat{T}_c(p_x \cdot (\mathbf{C}_{\mathbf{H}_0})_{l,l}) - N \right). \quad (34)$$

- 2) Using wideband constant amplitude modulation, each diagonal element of the conditional channel covariance matrix in the frequency domain, given all past transmitted and received symbols, is upper bounded by:

$$\lim_{k \rightarrow \infty} (\tilde{\mathcal{E}}_k)_{m,m} \leq \frac{1}{N} \frac{1}{1 + \frac{p_x}{2L} \left(\hat{T}_c(p_x) - N \right)}. \quad (35)$$

Equality in (35) is achieved if the channel energy is concentrated into a single tap (i.e., if $L = 1$).

- 3) Using narrowband (i.e., only a single frequency bin is active, using all of the OFDM symbol power, $r = 1$) constant amplitude modulation, the relevant diagonal element of the conditional channel covariance matrix in the frequency domain, given all past transmitted and received symbols, is given by:

$$\frac{1}{1 + \frac{p_x}{2} \left(\hat{T}_c(p_x) - N \right)}. \quad (36)$$

- 4) At the low SNR limit, define:

$$\hat{T}_{c_0} = \lim_{p_x \rightarrow 0} \hat{T}_c(p_x) = 2N \lim_{k \rightarrow \infty} D_k^\dagger D_k + N. \quad (37)$$

If this limit exists, it converges to the definition of coherence time used by Medard and Gallager [5].

- 5) For any $\lambda > 1$ the effective coherence time satisfies:

$$\hat{T}_c(p_x) / \lambda \leq \hat{T}_c(\lambda p_x) \leq \hat{T}_c(p_x). \quad (38)$$

- 6) If the prediction error is not zero, i.e.:

$$\mathcal{E}_\infty = \lim_{k \rightarrow \infty} \lim_{p_x \rightarrow \infty} \tilde{\mathcal{E}}_k = \lim_{k \rightarrow \infty} \mathbf{C}_{\tilde{\mathbf{H}}_0} [1 - D_k^\dagger \Delta_{k-1}^{-1} D_k] > 0 \quad (39)$$

then:

$$\lim_{p_x \rightarrow \infty} \hat{T}_c(p_x) = N. \quad (40)$$

Properties 1–3 are shown in the proof of Lemma 2.b in Section VII. They reflect the relation between the effective coherence time and the conditional channel distribution, which is the basis for the capacity bounds. They can also be used to argue the relationship between the coherence time and the effective coherence time. To particularize the results in a case where we have a direct and intuitive relation between capacity and coherence time, we consider a block fading channel [21]. In this model, the time axis is divided into blocks of length T_c/N , and the channel realization is fixed and independent within each block. This channel is nonstationary, and exhibits different statistics for each symbol (as an example, the first symbol of each block has correlation only with future symbols, while the last symbol of each block has correlation only with past symbols). As a reference, we assume that the k -th symbol is the middle symbol of a block and that the block length is an odd multiple of the symbol length N . It is easy to show⁵ that the conditional channel covariance matrix for the middle symbol of a block satisfies Properties 1–3 if we substitute $\hat{T}_c(p_x)$ with T_c , the block length. This similarity motivates the term *effective coherence time* as an extension of the coherence time definition to more general channel models.

Our simulations have shown that Properties 1–3 can also well approximate the conditional channel covariance matrix given all past transmitted and received symbols when using Gaussian distributed input signals, as long as the effective coherence time is large enough. The Gaussian modulation is of interest as it achieves the coherent channel capacity. So far, we have not been able to provide a formal proof of this, and this is the reason why we used truncated Gaussian modulation instead of Gaussian modulation for the capacity lower bound in Theorem 3. We demonstrate the accuracy of this approximation in the simulation results presented in Section V-C.

Property 4 is easily verified from (12). As long as \hat{T}_{c_0} exists, the effective coherence time can be seen as an extension of the coherence time defined in [5] to higher SNR. Note that

⁵For example by using $D_k = 1_d$ and $\Delta_{k-1} = 1_d 1_d^\dagger$ in (97) for the wideband case (with $\beta = 1$), and in (87) for the narrowband case, with $d = 0.5(T_c/N - 1)$.

the effective coherence time gives a good characterization of the capacity even for channels for which \hat{T}_{c_0} does not exist.

The major difference between any channel coherence time definition and the effective coherence time is that the latter is a function of the system SNR. This may seem a surprising characteristic of a coherence time, since intuitively one would expect the coherence time to characterize the channel, regardless of the system parameters. This dependence of the effective coherence time on SNR comes from the fact that at high SNR the system is more sensitive to changes in the channel, and hence will effectively see a shorter coherence time. Alternatively, one can say that a system with higher SNR requires better channel estimations. Thus, such a system will be able to use only channel measurements that present higher correlation with the present channel state, so that an increase in the SNR will reduce the number of useful measurements and consequently reduce the effective coherence time. Note that some engineers actually regard the SNR as a part of the channel and not as a system parameter (in the sense that in many cases the transmission power, just like the channel, is a limitation given to the system designer and not part of the system design).

Property 5 can be seen from (12) by considering the eigenvalue decomposition of the matrix $\Delta_{k-1} - D_k D_k^\dagger$. A more detailed description can be found in [22]. As reflected by Property 5, the effective coherence time is a nonincreasing function of the SNR, and for most channel models it decreases as the SNR increases.

Property 6 follows from Property 2 and deals with the high SNR extreme behavior of the effective coherence time. The prediction error is the conditional covariance of the channel given all past channel realizations (i.e., if the prediction error is zero then the channel is fully predictable). In regular channel models $\mathcal{E}_\infty^\infty$ is larger than zero [12], and represents the inherent uncertainty in the channel estimation given the full knowledge of its past. Property 6 shows that for regular channels the effective coherence time converges to N as the SNR grows to infinity (as we assume that the channel does not change within an OFDM symbol, the model cannot show a coherence time smaller than N).

As we limit the present analysis to large effective coherence times, for such regular channels the presented bounds will not be tight for very high SNR. The looseness of the bounds for very high SNR is in agreement with the results of Lapidot and Moser [12], which showed that the coherent channel capacity is not achievable in regular channels in the high SNR limit (the

capacity grows only doubly logarithmically with SNR).

Although not proved herein, the effective coherence time can significantly decrease even for irregular channels. As an example consider a Clarke's fading channel [23] with a doppler spread of 300 Hz and an OFDM symbol length of $66.7\mu\text{S}$ (which corresponds to an LTE mobile device operating at the 2GHz band and moving at a speed of 160KMH). The effective coherence time of this channel at 6dB SNR is 50 symbols, and the underspread assumption is well verified. However, at a very high SNR of 116dB the effective coherence time drops to only 10 symbols (at which point one might start questioning the underspread assumption).

B. Bound gap at finite SNR

We have discussed the bound tightness at the high and low SNR limits. In this subsection we discuss the gap between the bounds in finite SNR. We will show that in most cases, if the coherence time is large enough, this gap is not significant, and hence the bounds describe the channel capacity very well. For simplicity, the discussion in this subsection is limited to the case of a channel with zero mean ($E[\tilde{\mathbf{H}}_0] = 0$).

We start with the peak power constraint for $p_x \leq \frac{2}{\hat{T}_c(p_x)}$. Using $\log(1+x) \geq x - 0.5x^2$ the upper bound (16) is upper bounded by:

$$\text{UB}_{\text{low}}^{(\text{pk})}(p_x) \leq \frac{\frac{p_x^2}{2}}{1 + \frac{p_x}{2}(\hat{T}_c(p_x) - N)} \cdot \frac{\hat{T}_c(p_x) + \frac{1}{2}(\hat{T}_c(p_x) - N)^2 \cdot p_x}{1 + \frac{p_x}{2}(\hat{T}_c(p_x) - N)}. \quad (41)$$

For the lower bound we substitute in (23) the inequality⁶ $\log(1 + e^{-x}) \leq \log 2 - \frac{x}{2} + \frac{x^2}{8}$ which results in $C_{\text{QPSK}}(\rho) \geq \rho - \rho^2$. Substituting the inequality and $r = \beta = 1$ in the lower bound (22), we get:

$$\text{LB}_{\text{QPSK}}(p_x, 1, 1) \geq \frac{\frac{p_x^2}{2}(\hat{T}_c(p_x) - N)}{1 + \frac{p_x}{2}(\hat{T}_c(p_x) - N) + Np_x} - \frac{N\frac{p_x^4}{2}(\hat{T}_c(p_x) - N)^2}{\left(1 + \frac{p_x}{2}(\hat{T}_c(p_x) - N) + Np_x\right)^2}. \quad (42)$$

These upper and lower bounds are almost identical for very low SNR and high effective coherence time, and start diverging as the SNR increases. For the SNR range of interest, the largest gap appears at the highest SNR: $p_x = \frac{2}{\hat{T}_c(p_x)}$. Defining the factor $\kappa = \frac{\hat{T}_c(p_x) - N}{\hat{T}_c(p_x)}$ (which is

⁶Using: $\log(1 + e^{-x}) = \log 2 - \frac{x}{2} + \log \cosh(\frac{x}{2})$, and $\cosh(x) \leq e^{\frac{1}{2}x^2}$.

approximately 1 for large enough effective coherence time) we have $Np_x = 2 - 2\kappa$:

$$\text{UB}_{\text{low}}^{(\text{pk})} \left(\frac{2}{\hat{T}_c(p_x)} \right) \leq \frac{p_x}{1 + \kappa} \frac{1 + \kappa^2}{1 + \kappa}, \quad (43)$$

$$\text{LB}_{\text{QPSK}} \left(\frac{1}{\hat{T}_c(p_x)}, 1, 1 \right) \geq \frac{\kappa p_x}{3 - \kappa} - \frac{4(1 - \kappa)\kappa^2 p_x}{(3 - \kappa)^2}. \quad (44)$$

For example, if $p_x = \frac{2}{\hat{T}_c(p_x)}$ is satisfied for $\hat{T}_c(p_x) = 100N$ then $\kappa = 0.99$ and the gap between (43) and (44) is only 2.5%.

Next, for $\frac{2}{\hat{T}_c(p_x)} \leq p_x \leq \frac{1}{2}$ (and still for the peak power constraint), we turn to a numerically verified bound: For $\mathbf{u} \sim \mathcal{CN}(0, 1)$ and $\rho < 0.5$ one can verify that:

$$E [C_{\text{QPSK}}(\rho|\mathbf{u}|^2)] \geq 0.97E [\log(1 + \rho|\mathbf{u}|^2)]. \quad (45)$$

Using this inequality in (22) and substituting $r = N$, $\beta = 1$ we have:

$$\begin{aligned} \text{LB}_{\text{QPSK}}(p_x, N, 1) &= E \left[C_{\text{QPSK}} \left(\frac{\frac{p_x^2}{2} (\hat{T}_c(p_x) - N) |\mathbf{u}|^2}{\frac{p_x}{2} (\hat{T}_c(p_x) - N) + L(1 + p_x)} \right) \right] \\ &= E \left[C_{\text{QPSK}} \left(\frac{p_x |\mathbf{u}|^2}{1 + \frac{2L(1 + p_x)}{p_x (\hat{T}_c(p_x) - N)}} \right) \right] \\ &\geq 0.97E \left[\log \left(1 + \frac{p_x |\mathbf{u}|^2}{1 + \frac{2L(1 + p_x)}{p_x (\hat{T}_c(p_x) - N)}} \right) \right]. \end{aligned} \quad (46)$$

Comparing this bound with the coherent channel upper bound, (15), and recalling that $\tilde{\mathbf{h}}_{0,0} \sim \mathcal{CN}(0, \frac{1}{N})$ one can evaluate the gap between the bounds. Again this gap is largest at the high SNR limit, $p_x = 0.5$. Considering the example of $\hat{T}_c(0.5) = 100N$ and $N = 5L$, the gap between the bounds is at most 4%.

Thus, for the peak power constraint case the bounds are good if the effective coherence time is large enough for SNRs of up to $p_x \leq 0.5$. As stated in Section IV, for higher SNRs the bounds are less tight for the peak power case (due to the power decrease that is required to allow for the truncated Gaussian modulation to satisfy the peak power constraint). The bounds can be tight again only at much higher SNR where the power penalty term (which results in an asymptotic difference of 1.2 nats) becomes negligible (if the effective coherence time is still large enough).

In the case of the quadratic power constraint we already stated the bounds' tightness when the effective coherence time is large enough for the high SNR limit and for the low SNR limit if the limit \hat{T}_{c_0} exists. We next quantify the bounding gap for finite SNR.

In the low SNR regime the bounds can be quite loose, depending on the behavior of the effective coherence time as a function of the SNR. Typically, the bounds are less tight at the point where the two upper bounds intersect. If $\hat{T}_c(p_x) \gg N$ then the two upper bounds ((15) and (17)) intersect very close to $p_x = \frac{2}{(\alpha-1)\hat{T}_c(p_x)}$. Using the same derivation as in (42) and substituting $r = 1$ and $\beta = \alpha$, we can lower bound the QPSK lower bound by:

$$\text{LB}_{\text{QPSK}}(p_x, 1, \alpha) \geq \frac{\alpha \frac{p_x^2}{2} \left(\hat{T}_c(\alpha p_x) - N \right)}{1 + N\alpha p_x + \frac{\alpha p_x}{2} \left(\hat{T}_c(\alpha p_x) - N \right)} - \frac{N\alpha^3 \frac{p_x^4}{2} \left(\hat{T}_c(\alpha p_x) - N \right)^2}{\left(1 + N\alpha p_x + \frac{\alpha p_x}{2} \left(\hat{T}_c(\alpha p_x) - N \right) \right)^2}.$$

Redefining $\kappa = \frac{\hat{T}_c(\alpha p_x) - N}{\hat{T}_c(\alpha p_x)}$, and defining $v = \frac{\alpha}{\alpha-1} \frac{\hat{T}_c(\alpha p_x)}{\hat{T}_c(p_x)}$ we have $v = \alpha p_x \hat{T}_c(\alpha p_x)/2$, $N\alpha p_x = v(2 - 2\kappa)$ and

$$\text{LB}_{\text{QPSK}} \left(\frac{2}{(\alpha-1)\hat{T}_c(p_x)}, 1, \alpha \right) \geq \frac{v\kappa p_x}{1 + v(2 - \kappa)} - \frac{2v^3\kappa^2(2 - 2\kappa)p_x}{(1 + v(2 - \kappa))^2}$$

while the upper bounds intersect very close to:

$$\text{UB}_{\text{low}}^{(\text{qd})}(p_x) \approx \text{UB}_{\text{coh}}(p_x) \approx p_x \quad (47)$$

Nothing that for high enough effective coherence time $\kappa \approx 1$, the bounding gap is mostly determined by the term v . Using Property 5 from Subsection V-A, v is lower bounded by $v \geq \frac{1}{\alpha-1}$ and hence the bounds can differ by approximately a factor of α . However, typically, the effective coherence time will not change that fast. If the effective coherence time is approximately constant at the relevant SNRs ($\hat{T}_c(\alpha p_x) \approx \hat{T}_c(p_x)$) the ratio between the bounds will be approximately $\frac{2\alpha-1}{\alpha}$ (i.e., between 1.5 for $\alpha = 2$ and 2 for high α).

For higher SNR, the bounding gap becomes smaller. Inspecting the lower bound (26), we have two penalty terms compared to the upper bound (15). The first is a capacity penalty term (the second term on the right hand side), and the second is an SNR penalty term (at the denominator of (28) inside the log). As shown above, the bounds are tight for high enough SNR and high enough effective coherence time. Taking as a reference SNR of $p_x = 1$ ($\text{UB}_{\text{coh}}(p_x) > 0.6$ nats), setting $r = N$, $\beta = 1$, and $\eta = 0.005$ will result in a capacity penalty term of 0.005 nats. If the effective coherence time satisfies $\hat{T}_c(p_x) > 330L + N$ then the SNR penalty term will be less

than 0.3 dB, and the gap between the bounds will be less than 5%. For higher SNRs (satisfying the same condition on the effective coherence time) or for higher effective coherence times, the gap will be even smaller. Note however that for high enough SNRs the effective coherence time will typically decrease to a level that will not allow coherent communication, and the bounds will diverge.⁷

Numerical examples showing the bounds gap are shown in the next subsection.

C. Numerical example

In order to visually demonstrate the bounds derived in the previous sections, we evaluate them numerically for the auto-regressive channel model of the first order (AR1). This channel model is defined by a single parameter, γ , the channel forgetting factor, and is characterized by $C_{\mathbf{H}_i, \mathbf{H}_j} = \gamma^{|i-j|} C_{\mathbf{H}_0}$. The effective coherence time for this channel can be calculated using (12), and its low SNR limit is $\hat{T}_{c_0} = N(1 + \gamma^2)/(1 - \gamma^2)$. Throughout this section we assume $E[\mathbf{H}_k] = 0$, equal power taps (i.e., $(C_{\mathbf{H}_k})_{l,l} = 1/L$ for $0 \leq l < L$) and $N = 30$, $L = 5$.

Figure 1 shows the effective coherence time of the AR1 channel for different values of the channel forgetting factor. In all cases the effective coherence time reaches its limit, \hat{T}_{c_0} , at low enough SNR (for $\hat{T}_{c_0} \geq 50,000$ this convergence is not seen in the figure as it happens in lower SNR). For higher SNR the effective coherence time is a decreasing function of the SNR, until it reaches $\hat{T}_c(p_x) = N$ which is the lowest value measurable in our model.

Figure 2 shows the bounds on the capacity of the AR1 channel with quadratic power constraint when the quadratic constraint constant in (11) is set to $\alpha = 10$. The figure shows the capacity bounds when the channel forgetting factor is $\gamma = 0.9672$ ($\hat{T}_{c_0} = 900$). The channel capacity is upper bounded by the low SNR upper bound, $UB_{\text{low}}^{(\text{qd})}$ (17), which is effective for low SNRs, and by the coherent channel upper bound, UB_{coh} (51), which is effective in higher SNRs. In order to demonstrate the role of the different lower bounds we draw 3 lower bound curves. Two of the bounds are based on the QPSK bound (22). For low SNRs the tightest bound uses narrowband QPSK signaling: $LB_{\text{QPSK,nw}}^{(\text{qd})} = \max_{1 \leq \beta \leq \alpha} LB_{\text{QPSK}}(p_x, 1, \beta)$. For medium SNRs the tightest bound uses wideband QPSK signaling: $LB_{\text{QPSK,wd}}^{(\text{qd})} = \max_{1 \leq \beta \leq \alpha} LB_{\text{QPSK}}(p_x, N, \beta)$.

⁷The two lower bounds in this work assume a standard coherent communication scheme, i.e., the receiver estimates the channel based on past symbols, and then uses this estimate to detect the next symbol.

For high SNRs the tightest bound uses the truncated Gaussian signaling, (26): $LB_{TG,wd}^{(qd)} = \sup_{0 < \eta < 1} LB_{TG}(p_x, N, 1, \eta)$. Note that the upper bounds intersect at $p_x = -38.5$ dB. At this point the ratio between the upper bounds and lower bounds is 0.5. For higher and lower SNR the bounds get tighter, up to the point where, for high enough SNR, the effective coherence time decreases too much and the bounds diverge.

The figure also demonstrates the achievable rates using Gaussian signaling, and the accuracy of the approximation suggested in Section V-A. Using Montecarlo simulations, we evaluated the bound for proper complex Gaussian input distribution (using (79) with $\eta = 0$ and the conditional channel covariance matrix from (84)). The resulting rate is depicted in x-marks in Figure 2 (such an evaluation is of course feasible only for short coherence times). The figure also depicts (in a dashed line) the resulting rate when the conditional channel covariance matrix is approximated according to Properties 1–3 in Subsection V-A (which were derived for the case of constant amplitude modulation, and are used here for the approximation of the capacity in the case of Gaussian modulation). As it can be seen, at least for the plotted case, the suggested approximation is very accurate and the approximation error is negligible. The large number of simulations performed have supported this claim, and shown that the approximation accuracy is even better for channels with longer coherence times. Based on these observations we suggest that Gaussian signaling may lead to an even tighter lower bound.

Figure 3 depicts the combined capacity bounds for various values of the channel forgetting factor when the quadratic constraint constant is $\alpha = 2$. The upper bound (depicted by solid lines) is the bound given by Theorem 1. The lower bound (depicted by dashed lines) is the maximum of the bounds given by Theorems 2 and 3. The figure depicts the bounds for forgetting factors of $\gamma = 0.9851, 0.997, 0.9994$ (which correspond to $\hat{T}_{c_0} = 2, 000, 10, 000$, and $50, 000$ respectively).

As it can be seen, the bounds are good in most of the range. The bounds are least tight at SNR values in which the two upper bounds intersect. In these SNR the ratio between the upper and lower bounds is 1.89, 1.64 and 1.59 for $\hat{T}_{c_0} = 2, 000, 10, 000$ and $50, 000$ respectively, very close to the ratio predicted in Section V-B for the case of slowly changing $\hat{T}_c(p_x)$. For higher and lower SNR the bounds are much tighter. In particular, for high SNR all lower bounds are close to the coherent channel upper bound, but the lower bound is tighter for channels with higher effective coherence time.

The x-marks in the top right end of the lower bounds (for $\hat{T}_{c_0} = 2, 000$ and $\hat{T}_{c_0} = 10, 000$) mark

the point in which the effective coherence time dropped below $\hat{T}_c(p_x) = 2N = 60$. For higher SNR the lower bounds will not be tight. More importantly, for higher SNR our model will not represent well the physical channel, because the assumption that the channel impulse response does not change during one OFDM symbol can no longer be justified. For longer coherence times, this SNR threshold is higher and hence not shown in the figure. For any SNR below the x-marks, the bounds derived in the previous sections are good, and describe the channel capacity with high accuracy.

D. Comparison to known results

1) *Closest results:* The bounds which are most similar to the results presented here were derived by Sethuraman et al. [7] and by Durisi et al. [3]. Sethuraman et al. present an upper bound for the frequency-flat fading case ($N = 1$) with a peak power constraint. In this case the upper bound in [7] is tighter than the bounds in (18), but the difference is negligible for underspread channels ($\hat{T}_{c_0} \gg N$). The difference in the upper bound results from the relaxation $\log(1+x) \leq x$ in (55). This relaxation can be easily avoided, but the resulting bounds are much more complicated for $N > 1$, (in particular in the context of underspread channels and effective coherence time) while the difference between the bounds is negligible. Durisi et al. presented upper and lower bounds for frequency selective fading channels, but only for the case of a peak power constraint both in time and in frequency. Both works also presented results on low SNR capacity asymptote that will be discussed in the next subsection.

2) *Low SNR limit when \hat{T}_{c_0} exists:* In the low SNR limit, if the limit of the effective coherence time, \hat{T}_{c_0} , exists as the SNR goes to zero (37), the channel capacity is known and matches the results presented above both with quadratic power constraint and peak power constraint. For the quadratic power constraint [5], the capacity at the low SNR limit is $\alpha T_c \frac{E[x^4]}{2}$, which is exactly equal to the low SNR limit of the upper bound (19). Note that the low SNR limit of the lower bound (25) is lower by a factor of $(\hat{T}_{c_0} - N)/\hat{T}_{c_0}$ than the upper bound, which is negligible for $\hat{T}_{c_0} \gg N$. (The main reason for this difference between the upper and lower bounds is the lower bounding transmission scheme which assumes that each OFDM symbol (N samples) is decoded using channel estimation based only on past symbols.)

For the peak power constraint flat fading case ($N = 1$), the capacity is $p_x^2(\hat{T}_{c_0} - 1)/2$ [7], which matches the results presented here (again up to a negligible factor of $(\hat{T}_{c_0} - N)/\hat{T}_{c_0}$ as

discussed above). Note that for the peak power constraint we did not analyze the case where the average power constraint is lower than the peak power constraint.

For $N > 1$ the same work shows that the low SNR capacity asymptote for the delay separable underspread stationary channel is also given by $p_x^2(\hat{T}_{c_0} - 1)/2$. Under the assumption $\hat{T}_{c_0} \gg N$ this channel is approximately equal to the OFDM channel considered here, and the low SNR asymptotes also match. This result also matches the low SNR asymptotes presented in [3] for the case of peak power constraint only in time. It is interesting to note that for this case the peak power constraint used in [7], [3] is stricter than the peak power constraint used here. The power constraint in this work limits the energy of a single OFDM symbol. Translating to the time domain the constraint is applied to the average energy of groups of N samples. On the other hand the constraint in [7] applies for each time domain sample. Interestingly, the bounds are almost identical, even though [7] clearly shows that relaxing the peak power constraint results in a higher capacity for the same average transmission power. Comparing these two works one can conclude that the relaxation of the peak power constraint is effective only if it is applicable for time periods which are at least on the order of the channel coherence time. Allowing signal “peakiness” which must be averaged over periods (N) which are significantly shorter than the channel coherence time (\hat{T}_{c_0}) is not sufficient to increase the channel capacity.

Since the effective coherence time can be interpreted as the block length of a block fading model that achieves the same capacity, our results naturally match results that were derived for the block fading model (e.g., [6], [8]).

3) Analysis for a given channel estimation error: Several works (e.g. [18], [24]) consider the effect of channel estimation error with a given variance on the channel capacity. These works have been the basis for the lower bounds presented here, but they miss the effect of the transmitted signal on the ability to estimate the channel ([18] considers also an estimation from an out-of-band pilot signal). In this sense, one can say that the QPSK lower bound presented above (Theorem 2) is the most straightforward part of the work. It combines an extension of the bounds in [18], [24] to the OFDM model, the channel estimation scheme of [25], [26] which allows to estimate the channel using all (relevant) past transmitted and received symbols, and results on channel estimation errors for constant amplitude modulations. Perhaps the most important contribution of Theorem 2 is the derivation of the bound in a way that shows the role of the effective coherence time. Note that the truncated Gaussian lower bound (Theorem 3) is

more complicated as the modulation is not constant amplitude, and the upper bounds are more complicated as they cannot assume any estimation scheme.

4) *High SNR limit:* As detailed above, for regular channels our analysis holds only up to some finite SNR. The results show that the coherent channel upper bound is achievable as long as the effective coherence time is large enough. This is in agreement with known results [12], [13], [14] showing that in the very high SNR limit the coherent channel upper bound is not achievable. Our analysis cannot predict the actual double logarithmic behavior of the capacity, as it happens at too high SNR where the assumption on the value of the effective coherence time is not valid.

For irregular channels, although not proven, the effective coherence time also typically decreases as SNR increases (although at lower rate). Hence, our analysis is limited in SNR even for irregular channels.

The modeling problem at high SNR is also discussed by Durisi et al. [27], where they focus on the characterization of the range of SNRs in which the channel discretization is reliable. Note that their approach is quite different from the one taken above. In [27], the analysis starts from a continuous time channel, and tries to analyze all modeling errors in the discretization process. In the analysis above we study the behavior of a discrete time OFDM model without assuming any discretization imperfections. Yet, we reach the same conclusion, even the discrete time OFDM model reveals the high SNR model limit.

5) *Dependence of coherence time on SNR:* Few works analyze the dependence between the SNR and the coherence time. In these works the coherence time characterizes the channel (and does not change for each channel). For example Zheng et al. [28] consider a block-fading channel, where the channel remains constant for a block of l symbols, before changing to an independent realizations. Chen and Veeravalli [29] consider a block stationary channel where the fading is constant across a block of length T and changes in a stationary manner between blocks.

In both cases the analysis considers a set of channels, and tries to characterize relations between the channel coherence time and the system SNR that will result in a certain capacity behavior. Chen and Veeravalli show ([29] equation (21)) that two systems with identical block correlation have the same capacity if the product of their peak powers and block lengths ($T \cdot \text{SNR}$) is identical.

Zheng et al. analyze the limit capacity of a set of channels with increasing coherence time l

and decreasing SNRs. They show that different relations between coherence time and SNR of the form $l = \text{SNR}^\alpha$ achieve different capacity behavior between the coherent and noncoherent scheme.

Our analysis is completely different. We consider a (single) specific channel, and show that its effective coherence time is a function of the system SNR.

VI. CHANNEL CAPACITY AND BOUNDS

In this section we analyze the channel capacity and give the information-theoretical part of the theorem proofs.

In general, even if the input symbols are independent, the output symbols are dependent due to the channel memory. Therefore, we need to evaluate the capacity over the entire transmission, defined as:

$$C = \lim_{n \rightarrow \infty} \sup_{\Pr(\tilde{\mathbf{X}}_0^n)} \frac{1}{N(n+1)} I(\tilde{\mathbf{X}}_0^n; \tilde{\mathbf{Y}}_0^n). \quad (48)$$

Note that the ergodic requirement in Assumption IV guarantees the achievability of the capacity (see for example [12], [6]).

Most of the bounds derived in this section depend on the conditional distribution of the channel given the past transmitted and received symbols. This distribution will be analyzed in detail in the next section. For this section, it is enough to state that given the past symbols, the channel has a complex Gaussian distribution $\tilde{\mathbf{H}}_k | \tilde{\mathbf{X}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1} \sim \mathcal{CN}(\tilde{\boldsymbol{\mu}}_k, \tilde{\boldsymbol{\Sigma}}_k)$. Lemma 3.a also uses the distribution of the channel mean:

$$\tilde{\boldsymbol{\mu}}_k | \tilde{\mathbf{X}}_0^{k-1} \sim \mathcal{CN}\left(E[\tilde{\mathbf{H}}_k], \mathbf{C}_{\tilde{\mathbf{H}}_k} - \tilde{\boldsymbol{\Sigma}}_k\right) \quad (49)$$

A. Coherent channel upper bound

The first upper bound we use is the well known channel capacity when the channel is known to the receiver (coherent channel) with only average power constraint:

$$\begin{aligned} C &\leq \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} \sup_{\Pr(\tilde{\mathbf{X}}_0^n)} I(\tilde{\mathbf{X}}_0^n; \tilde{\mathbf{Y}}_0^n, \tilde{\mathbf{H}}_0^n) \\ &= \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} \sum_{k=0}^n \sup_{\Pr(\tilde{\mathbf{X}}_k)} I(\tilde{\mathbf{X}}_k; \tilde{\mathbf{Y}}_k, \tilde{\mathbf{H}}_k). \\ &= \frac{1}{N} \sup_{\Pr(\tilde{\mathbf{X}}_0)} I(\tilde{\mathbf{X}}_0; \tilde{\mathbf{Y}}_0, \tilde{\mathbf{H}}_0). \end{aligned} \quad (50)$$

Using Assumptions IV and VI, this bound is maximized when the input signal has Gaussian distribution with $E[\tilde{\mathbf{X}}_k] = 0$, and $C_{\tilde{\mathbf{X}}_k} = p_x \cdot \mathbf{I}$ [30]. The maximal mutual information is given by:

Result 1: The channel capacity is upper bounded by

$$C \leq \text{UB}_{\text{coh}}(p_x) = E_{\tilde{\mathbf{H}}_0} \left[\log \left(1 + N p_x |\tilde{\mathbf{h}}_{0,0}|^2 \right) \right]. \quad (51)$$

The bound in (51) is the capacity of the channel with a constraint only on the average power. As both types of power constraints analyzed here must satisfy $E[\mathbf{p}_k] \leq N p_x$, this bound holds for both types of constraints. As in the case of the block fading model [21], we will show that for large enough effective coherence times and SNR this bound is tight, and the channel capacity does not degrade due to the lack of channel knowledge.

B. Low SNR upper bound

The channel capacity is upper bounded by the following lemma:

Lemma 1.a: The capacity of a channel with a peak power constraint is upper bounded by:

$$C \leq \lim_{k \rightarrow \infty} \left\{ p_x \left[1 + N \left| E[\tilde{\mathbf{h}}_{0,0}] \right|^2 \right] - \frac{1}{N} \log \left(1 + N^2 p_x \min_m \inf_{\tilde{\mathbf{X}}_0^{k-1} \in \mathcal{I}(N p_x \mathbf{I})} (\tilde{\boldsymbol{\epsilon}}_k)_{m,m} \right) \right\}, \quad (52)$$

and the capacity of a channel with a quadratic power constraint is upper bounded by:

$$C \leq \lim_{k \rightarrow \infty} \left\{ N p_x \left| E[\tilde{\mathbf{h}}_{0,0}] \right|^2 + \sqrt{\alpha} p_x \sqrt{\sup_{\Pr(\mathcal{P}_{k-1})} E \left[\left\{ 1 - N \min_m \inf_{\tilde{\mathbf{X}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\boldsymbol{\epsilon}}_k)_{m,m} \right\}^2 \right]} + \frac{1}{2} \alpha N p_x^2 \right\}, \quad (53)$$

where $\mathcal{I}(\mathcal{P}_k) = \{\tilde{\mathbf{X}}_0^k : \tilde{\mathbf{X}}_i^\dagger \tilde{\mathbf{X}}_i = \mathbf{p}_i, i = 0, \dots, k\}$ is the set of all input symbols that correspond to the power matrix \mathcal{P}_k .

Proof of Lemma 1.a: We start with the inequality:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{N(n+1)} I(\tilde{\mathbf{X}}_0^n; \tilde{\mathbf{Y}}_0^n) &= \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} \sum_{k=0}^n I(\tilde{\mathbf{Y}}_k; \tilde{\mathbf{X}}_0^n | \tilde{\mathbf{Y}}_0^{k-1}) \\
&= \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} \sum_{k=0}^n \left\{ h(\tilde{\mathbf{Y}}_k | \tilde{\mathbf{Y}}_0^{k-1}) - h(\tilde{\mathbf{Y}}_k | \tilde{\mathbf{X}}_0^n, \tilde{\mathbf{Y}}_0^{k-1}) \right\} \\
&\leq \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} \sum_{k=0}^n \left\{ h(\tilde{\mathbf{Y}}_k) - h(\tilde{\mathbf{Y}}_k | \tilde{\mathbf{X}}_0^n, \tilde{\mathbf{Y}}_0^{k-1}) \right\}. \quad (54)
\end{aligned}$$

where $h(\cdot)$ is the differential entropy. The first term in the last line of (54) can be upper bounded by the maximal entropy of a random vector with a given covariance matrix. The resulting entropy is [31]:

$$h(\tilde{\mathbf{Y}}_k) \leq \log(\|\mathbf{C}_{\tilde{\mathbf{Y}}_k}\|) + N \log(\pi e) \leq \text{Tr}\left(E[\tilde{\mathbf{Y}}_k \tilde{\mathbf{Y}}_k^\dagger] - \mathbf{I}\right) + N \log(\pi e). \quad (55)$$

where the last inequality uses $\log(\|\mathbf{I} + \mathbf{A}\|) \leq \text{Tr}(\mathbf{A})$ which holds for any positive semi-definite matrix \mathbf{A} .⁸ Note that this inequality is tight for low SNR if the transmitted signal has zero mean. Substituting the covariance matrix:

$$E[\tilde{\mathbf{Y}}_k \tilde{\mathbf{Y}}_k^\dagger] = N \cdot E\left[\text{diag}(\tilde{\mathbf{X}}_k) \tilde{\mathbf{H}}_k \tilde{\mathbf{H}}_k^\dagger \text{diag}(\tilde{\mathbf{X}}_k)^\dagger\right] + \mathbf{I}, \quad (56)$$

we get:

$$\begin{aligned}
h(\tilde{\mathbf{Y}}_k) - N \log(\pi e) &\leq N \cdot \text{Tr}\left(E[\mathcal{S}_{\mathbf{X}_k}] \left[\mathbf{C}_{\tilde{\mathbf{H}}_0} + E[\tilde{\mathbf{H}}_0] E[\tilde{\mathbf{H}}_0]^\dagger\right]\right) \\
&= E[\tilde{\mathbf{X}}_k^\dagger \tilde{\mathbf{X}}_k] \left[1 + N \left|E[\tilde{\mathbf{h}}_{0,0}]\right|^2\right] \quad (57)
\end{aligned}$$

where the first line uses the rotation property of the trace and the spectrum definition, (2). The second line uses the fact that $\mathcal{S}_{\mathbf{X}_k}$ is a diagonal matrix, while all elements on the diagonal of the second term are equal. We also use the normalization $\text{Tr}[\mathbf{C}_{\tilde{\mathbf{H}}_0}] = 1$, which results in $(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{m,m} = 1/N$.

Turning to the second entropy in the last line of (54), we note that the conditional distribution of the output given the input is Gaussian. We also observe that:

$$\mathbf{C}_{\tilde{\mathbf{Y}}_k | \tilde{\mathbf{X}}_0^n, \tilde{\mathbf{Y}}_0^{k-1}} = N \cdot \text{diag}(\tilde{\mathbf{X}}_k) \tilde{\mathbf{E}}_k \text{diag}(\tilde{\mathbf{X}}_k)^\dagger + \mathbf{I}, \quad (58)$$

⁸Using $\|\mathbf{I} + \mathbf{A}\| = \prod_{i=0}^{N-1} (1 + \lambda_i)$, and $\text{Tr}(\mathbf{A}) = \sum_{i=0}^{N-1} \lambda_i$, where $\lambda_0, \dots, \lambda_{N-1}$ are the eigenvalues of the matrix \mathbf{A} , and the inequality $\log(1+x) \leq x$.

where $\tilde{\mathbf{E}}_k$ is the conditional channel covariance matrix given the past transmitted and received symbols. Using (9), the entropy is written as:

$$\begin{aligned}
h(\tilde{\mathbf{Y}}_k | \tilde{\mathbf{X}}_0^n, \tilde{\mathbf{Y}}_0^{k-1}) - N \log(\pi e) &= E_{\tilde{\mathbf{X}}_0^n, \tilde{\mathbf{Y}}_0^{k-1}} \left[\log \left(\left\| \mathbf{C}_{\tilde{\mathbf{Y}}_k | \tilde{\mathbf{X}}_0^n, \tilde{\mathbf{Y}}_0^{k-1}} \right\| \right) \right] \\
&= E_{\mathcal{P}_n} \left[E_{\tilde{\mathbf{X}}_0^n | \mathcal{P}_n} \left[\log \left(\left\| N \text{diag}(\tilde{\mathbf{X}}_k) \tilde{\mathbf{E}}_k \text{diag}(\tilde{\mathbf{X}}_k)^\dagger + \mathbf{I} \right\| \right) \right] \right] \\
&= E_{\mathcal{P}_k} \left[E_{\tilde{\mathbf{X}}_0^k | \mathcal{P}_k} \left[\log \left(\left\| N \tilde{\mathbf{E}}_k \mathcal{S}_{\mathbf{X}_k} + \mathbf{I} \right\| \right) \right] \right] \\
&\geq E_{\mathcal{P}_k} \left[\inf_{\tilde{\mathbf{X}}_0^k \in \mathcal{I}(\mathcal{P}_k)} \log \left(\left\| N \tilde{\mathbf{E}}_k \mathcal{S}_{\mathbf{X}_k} + \mathbf{I} \right\| \right) \right], \tag{59}
\end{aligned}$$

where we use the rotation property $\|\mathbf{I} + \mathbf{AB}\| = \|\mathbf{I} + \mathbf{BA}\|$.

Now, we observe that since $\mathcal{S}_{\mathbf{X}_k}$ is diagonal and nonnegative,

$$\text{Tr} \left(\tilde{\mathbf{E}}_k \mathcal{S}_{\mathbf{X}_k} \right) \geq \min_m (\tilde{\mathbf{E}}_k)_{m,m} \text{Tr} (\mathcal{S}_{\mathbf{X}_k}). \tag{60}$$

Using also the inequality $\log \|\mathbf{I} + \mathbf{A}\| \geq \log(1 + \text{Tr}(\mathbf{A}))$ which holds for any positive semi-definite matrix \mathbf{A} ,⁹ the term inside the expectation in the last row of (59) is lower bounded by:

$$\begin{aligned}
\inf_{\tilde{\mathbf{X}}_0^k \in \mathcal{I}(\mathcal{P}_k)} \log \left(\left\| N \tilde{\mathbf{E}}_k \mathcal{S}_{\mathbf{X}_k} + \mathbf{I} \right\| \right) &\geq \inf_{\tilde{\mathbf{X}}_0^k \in \mathcal{I}(\mathcal{P}_k)} \log \left(1 + N \text{Tr} \left(\tilde{\mathbf{E}}_k \mathcal{S}_{\mathbf{X}_k} \right) \right) \\
&\geq \log \left(1 + N \mathbf{p}_k \min_m \inf_{\tilde{\mathbf{X}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\mathbf{E}}_k)_{m,m} \right). \tag{61}
\end{aligned}$$

Note that this lower bound is achievable using a transmission spectrum that concentrates all of the power on the frequency bin that has minimal estimation error.

Substituting (61) and (57) in (54) results in:

$$\begin{aligned}
C &\leq \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} \sum_{k=0}^n \sup_{\text{Pr}(\mathcal{P}_k)} \left\{ E[\mathbf{p}_k] \left[1 + N \left| E[\tilde{\mathbf{h}}_{0,0}] \right|^2 \right] \right. \\
&\quad \left. - E \left[\log \left(1 + N \mathbf{p}_k \min_m \inf_{\tilde{\mathbf{X}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\mathbf{E}}_k)_{m,m} \right) \right] \right\}. \tag{62}
\end{aligned}$$

Noting also that the resulting quantity is monotonically nondecreasing in k , the capacity is upper bounded by:

$$\begin{aligned}
C &\leq \frac{1}{N} \lim_{k \rightarrow \infty} \sup_{\text{Pr}(\mathcal{P}_k)} \left\{ E[\mathbf{p}_k] \left[1 + N \left| E[\tilde{\mathbf{h}}_{0,0}] \right|^2 \right] \right. \\
&\quad \left. - E \left[\log \left(1 + N \mathbf{p}_k \min_m \inf_{\tilde{\mathbf{X}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\mathbf{E}}_k)_{m,m} \right) \right] \right\}. \tag{63}
\end{aligned}$$

⁹Using $\sum_i \log(1 + \lambda_i) \geq \log(1 + \sum_i \lambda_i)$ since $\lambda_i \geq 0$.

Subject to the peak power constraint, for every k the bound is maximized if $\mathbf{p}_k = Np_x$ (as \mathbf{p}_k appears inside the log in the second (negative) term). Substituting this optimal transmission power in (63) results in (52) and proves the first part of the lemma.

In order to derive the bound for the quadratic power constraint we use the inequality $\log(1 + x) \geq x - \frac{1}{2}x^2$ and simplify (63) to:

$$C \leq \frac{1}{N} \lim_{k \rightarrow \infty} \sup_{\Pr(\mathcal{P}_k)} \left\{ E[\mathbf{p}_k] \left[1 + N \left| E[\tilde{\mathbf{h}}_{0,0}] \right|^2 \right] - E \left[N \mathbf{p}_k \min_m \inf_{\tilde{\mathbf{x}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\mathbf{e}}_k)_{m,m} - \frac{1}{2} N^2 \mathbf{p}_k^2 \left\{ \min_m \inf_{\tilde{\mathbf{x}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\mathbf{e}}_k)_{m,m} \right\}^2 \right] \right\}. \quad (64)$$

Using $(\tilde{\mathbf{e}}_k)_{m,m} \leq (C_{\tilde{\mathbf{H}}_0})_{m,m} = 1/N$, we rewrite (64) as:

$$C \leq \frac{1}{N} \lim_{k \rightarrow \infty} \sup_{\Pr(\mathcal{P}_k)} \left\{ N E[\mathbf{p}_k] \left| E[\tilde{\mathbf{h}}_{0,0}] \right|^2 + E \left[\mathbf{p}_k \left(1 - N \min_m \inf_{\tilde{\mathbf{x}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\mathbf{e}}_k)_{m,m} \right) + \frac{1}{2} \mathbf{p}_k^2 \right] \right\}, \quad (65)$$

and use the Cauchy-Schwartz inequality and the power constraint, (11), to prove (53). \blacksquare

The exact values of the bounds depend on the conditional channel distribution, through $\min_m \inf_{\tilde{\mathbf{x}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\mathbf{e}}_k)_{m,m}$. This distribution is analyzed in the next section.

C. QPSK lower bound

For the lower bound we need to select an input distribution. The first lower bound derived herein is based on QPSK modulation. This bound applies for both power constraints. For the quadratic power constraint, the bound is especially significant at the low SNR regime, where it is crucial to use a low fourth moment. The QPSK modulation uses constant amplitude and hence minimizes the ratio $\frac{E[|\tilde{\mathbf{x}}_{k,m}|^4]}{E^2[|\tilde{\mathbf{x}}_{k,m}|^2]}$.

The input symbol is given by:

$$\tilde{\mathbf{X}}_k = \sqrt{\beta \frac{N}{r}} \mathbf{g} \tilde{\mathbf{B}}_k, \quad (66)$$

where $\tilde{\mathbf{B}}_k = [\tilde{\mathbf{b}}_{k,0}, \dots, \tilde{\mathbf{b}}_{k,r-1}, 0, \dots, 0]$, r ($1 \leq r \leq N$) is the number of active frequency bins, $\{\tilde{\mathbf{b}}_{k,m}\}$ is the sequence of iid QPSK data, and $\tilde{\mathbf{b}}_{k,m} \in \{\pm 1, \pm j\}$ with equal probability. On the other hand, $\mathbf{g} \in \{0, 1\}$ is a binary random variable, statistically independent from $\tilde{\mathbf{b}}_{k,0}, \dots, \tilde{\mathbf{b}}_{k,r-1}$,

that determines whether the system will transmit or not (i.e., a single variable that affects all transmitted symbols). The transmission probability is $\Pr(\mathbf{g} = 1) = \frac{1}{\beta}$. This distribution is convenient for analysis, although it is not reasonable for practical systems (one cannot consider a practical system that has a positive probability not to transmit any symbol at any time). A practical system can achieve the same performance by transmitting in blocks with large gaps between the blocks, as long as the block length is significantly larger than the channel coherence time.

Lemma 2.a: The capacity of a channel with a quadratic power constraint is lower bounded by:

$$C \geq \lim_{k \rightarrow \infty} \max_{r \leq N} \max_{1 \leq \beta \leq \alpha} \frac{1}{N\beta} \sum_{m=0}^{r-1} E \left[C_{\text{QPSK}} \left(\frac{\beta \frac{N^2}{r} p_x |\tilde{\boldsymbol{\mu}}_{k,m}|^2}{1 + \beta \frac{N^2}{r} p_x (\tilde{\boldsymbol{\epsilon}}_k)_{m,m}} \right) \middle| \mathbf{g} = 1 \right], \quad (67)$$

where $C_{\text{QPSK}}(\rho)$, defined in (23), is the achievable mutual information for QPSK transmission over AWGN channel with SNR equal to ρ (see for example [20], which also gives some useful bounds). The capacity of a channel with a peak power constraint is lower bounded by:

$$C \geq \lim_{k \rightarrow \infty} \max_{r \leq N} \frac{1}{N} \sum_{m=0}^{r-1} E \left[C_{\text{QPSK}} \left(\frac{\frac{N^2}{r} p_x |\tilde{\boldsymbol{\mu}}_{k,m}|^2}{1 + \frac{N^2}{r} p_x (\tilde{\boldsymbol{\epsilon}}_k)_{m,m}} \right) \middle| \mathbf{g} = 1 \right]. \quad (68)$$

Proof of Lemma 2.a: As \mathbf{g} is constant throughout the transmission, we can safely assume that after long enough time the receiver can decode \mathbf{g} with no error. Thus, we can write:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} I(\tilde{\mathbf{X}}_0^n; \tilde{\mathbf{Y}}_0^n) &= \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} I(\tilde{\mathbf{X}}_0^n; \tilde{\mathbf{Y}}_0^n | \mathbf{g}) \\ &= \frac{1}{\beta} \lim_{n \rightarrow \infty} \frac{1}{N(n+1)} I(\tilde{\mathbf{B}}_0^n; \tilde{\mathbf{Y}}_0^n | \mathbf{g} = 1). \end{aligned} \quad (69)$$

Given $\mathbf{g} = 1$ the transmitted symbols are statistically independent. We use the following equality:

$$\begin{aligned} I(\tilde{\mathbf{B}}_0^n; \tilde{\mathbf{Y}}_0^n | \mathbf{g} = 1) &= \sum_{k=0}^n I(\tilde{\mathbf{B}}_k; \tilde{\mathbf{Y}}_k | \mathbf{g} = 1, \tilde{\mathbf{B}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1}, \tilde{\mathbf{Y}}_{k+1}^n) \\ &\quad + \sum_{k=0}^n I(\tilde{\mathbf{B}}_k; \tilde{\mathbf{Y}}_0^{k-1} | \mathbf{g} = 1, \tilde{\mathbf{B}}_0^{k-1}, \tilde{\mathbf{Y}}_{k+1}^n) \\ &\quad + \sum_{k=0}^n I(\tilde{\mathbf{B}}_k; \tilde{\mathbf{Y}}_{k+1}^n | \mathbf{g} = 1, \tilde{\mathbf{B}}_0^{k-1}) \end{aligned} \quad (70)$$

The second and third terms on the right hand side of this equation correspond to the mutual information between an input symbol and the past or future output symbols. As these terms do

not use the k -th symbol output, they both vanish when the input symbols are iid. The first term can be lower bounded by:

$$\begin{aligned}
I(\tilde{\mathbf{B}}_0^n; \tilde{\mathbf{Y}}_0^n | \mathbf{g} = 1) &= \sum_{k=0}^n I(\tilde{\mathbf{B}}_k; \tilde{\mathbf{Y}}_k | \mathbf{g} = 1, \tilde{\mathbf{B}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1}, \tilde{\mathbf{Y}}_{k+1}^n) \\
&\geq \sum_{k=0}^n I(\tilde{\mathbf{B}}_k; \tilde{\mathbf{Y}}_k | \mathbf{g} = 1, \tilde{\mathbf{B}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1}) \\
&\geq \sum_{k=0}^n \sum_{m=0}^{r-1} I(\tilde{\mathbf{b}}_{k,m}; \tilde{\mathbf{y}}_{k,m} | \mathbf{g} = 1, \tilde{\mathbf{B}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1}), \tag{71}
\end{aligned}$$

where the second line uses the statistical independence of $\tilde{\mathbf{Y}}_{k+1}^n$ and $\tilde{\mathbf{B}}_k$, and the third line uses the statistical independence between the elements of $\tilde{\mathbf{B}}_k$. Substituting (71) in (69), and noting again that the term inside the summation is monotonic in k , results in the lower bound on the capacity:

$$C \geq \lim_{k \rightarrow \infty} \frac{1}{N\beta} \sum_{m=0}^{r-1} I(\tilde{\mathbf{b}}_{k,m}; \tilde{\mathbf{y}}_{k,m} | \tilde{\mathbf{B}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1}, \mathbf{g} = 1). \tag{72}$$

Note that the past symbols affect the distribution of the current symbol only through the conditional distribution of the channel. Thus, in this lower bound, the decoding of correlated output symbols is replaced by the decoding of uncorrelated symbols with channel state information (CSI). This CSI is produced by optimal channel estimation based on past input and output symbols. A coding scheme that allows to decode the previous symbols and use them for channel estimation for the next symbol is shown in [25], [26].

Focusing on the m -th frequency bin of the k -th symbol, given $\mathbf{g} = 1$, the channel output is:

$$\tilde{\mathbf{y}}_{k,m} = \sqrt{\frac{\beta}{r}} p_x N \tilde{\boldsymbol{\mu}}_{k,m} \tilde{\mathbf{b}}_{k,m} + \tilde{\mathbf{w}}_{k,m} + \tilde{\mathbf{v}}_{k,m}. \tag{73}$$

where $\tilde{\mathbf{w}}_{k,m}$ is the m -th element of the DFT of the Complex Gaussian noise, and $\tilde{\mathbf{v}}_{k,m} = \sqrt{\frac{\beta}{r}} p_x N (\tilde{\mathbf{h}}_{k,m} - \tilde{\boldsymbol{\mu}}_{k,m}) \tilde{\mathbf{b}}_{k,m}$ is the interference term due to channel estimation error. Since the estimation error is a circularly symmetric Gaussian random variable and the data has constant amplitude, the interference term is also Gaussian and is statistically independent from the data ($\tilde{\mathbf{b}}_{k,m}$). Assuming optimal channel estimation, the resulting channel is equivalent to a coherent flat fading channel with additive independent complex Gaussian noise samples of variance $1 + \beta \frac{N^2}{r} p_x (\tilde{\boldsymbol{\Sigma}}_k)_{m,m}$. Evaluating the mutual information of this channel for QPSK modulation and multiplying it by the transmission probability $(1/\beta)$ leads to (67) and (68), depending on the allowed range of the parameter β . ■

D. Truncated Gaussian lower bound

For high SNR, the QPSK bound will not be tight (most obviously as it is limited to 2 bits per frequency bin). The most intuitive alternative is a complex Gaussian input distribution. Such a distribution seems very close to optimal (based on a large set of simulations). However, we were unable to produce an appropriate bound on the resulting conditional channel variance, and therefore turned to a slightly modified distribution.

Let \mathbf{u} be a proper complex Gaussian distribution, $\mathbf{u} \sim \mathcal{CN}(0, 1)$. We denote hereafter by *truncated Gaussian distribution* with parameters η and ξ the distribution of \mathbf{u} given : $\eta \leq |\mathbf{u}|^2 \leq \xi$. In mathematical terms, we say that $\mathbf{z} \sim \mathcal{TCN}(\eta, \xi)$ if for any function $f(\cdot)$ we have:

$$E[f(\mathbf{z})] = E\left[f(\mathbf{u}) \middle| \eta \leq |\mathbf{u}|^2 \leq \xi\right]. \quad (74)$$

Note that $|\mathbf{u}|^2$ has an exponential distribution with parameter 1. Define:

$$\theta = \Pr(\eta \leq |\mathbf{u}|^2 \leq \xi) = e^{-\eta} - e^{-\xi}. \quad (75)$$

The second moment of the truncated Gaussian distribution is:

$$p_z = \frac{1}{\theta} \int_{\eta}^{\xi} \chi e^{-\chi} d\chi = 1 + \frac{\eta e^{-\eta} - \xi e^{-\xi}}{e^{-\eta} - e^{-\xi}}. \quad (76)$$

The Fourth moment of the truncated Gaussian distribution is:

$$E[|\mathbf{z}|^4] = \frac{1}{\theta} \int_{\eta}^{\xi} \chi^2 e^{-\chi} d\chi = 2 + \frac{(\eta^2 + 2\eta)e^{-\eta} - (\xi^2 + 2\xi)e^{-\xi}}{e^{-\eta} - e^{-\xi}} \quad (77)$$

Using the truncated Gaussian distribution, we define the input symbol as:

$$\tilde{\mathbf{X}}_k = \frac{c}{\sqrt{p_z}} \mathbf{g} \tilde{\mathbf{Z}}_k, \quad (78)$$

where c is chosen to satisfy the power constraint, \mathbf{g} is defined as in the QPSK lower bound, $\tilde{\mathbf{Z}}_k = [\tilde{\mathbf{z}}_{k,0} \dots, \tilde{\mathbf{z}}_{k,r-1}, 0, \dots, 0]$, $1 \leq r \leq N$ is the number of active frequency bins, and $\{\tilde{\mathbf{z}}_{k,m}\}$ is the sequence of iid input random variables with truncated Gaussian distribution $\tilde{\mathbf{z}}_{k,m} \sim \mathcal{TCN}(\eta, \xi)$.

Lemma 3.a: If $E[\tilde{\mathbf{h}}_{0,m}] = 0$, the capacity of a channel with a quadratic power constraint (and $\alpha > 1$) is lower bounded by:

$$\begin{aligned} C \geq & \lim_{k \rightarrow \infty} \max_{r \leq N} \sup_{\eta > 0} \max_{1 \leq \beta \leq \frac{\alpha r(1+\eta)^2}{1+r(1+\eta)^2}} -\frac{r\eta}{N\beta} \\ & + \frac{1}{N\beta} \sum_{m=0}^{r-1} \log \left(1 + \frac{\beta \frac{N^2}{r} p_x \left(\frac{1}{N} - E\left[(\tilde{\mathbf{e}}_k)_{m,m} \middle| \mathbf{g} = 1\right]\right) |\mathbf{u}|^2}{1 + \beta \frac{N^2}{r} p_x E\left[(\tilde{\mathbf{e}}_k)_{m,m} \middle| \mathbf{g} = 1\right]} \right) \end{aligned} \quad (79)$$

and the capacity of a channel with a peak power constraint is lower bounded by:

$$C \geq \lim_{k \rightarrow \infty} \max_{r \leq N} \sup_{0 < \eta < \xi} \frac{1}{N} \sum_{m=0}^{r-1} \log \left(1 + \frac{\frac{N^2 p_x p_z}{r \xi} \left(\frac{1}{N} - E \left[(\tilde{\mathbf{E}}_k)_{m,m} \middle| \mathbf{g} = 1 \right] \right) |\mathbf{u}|^2}{1 + \frac{N^2 p_x p_z}{r \xi} E \left[(\tilde{\mathbf{E}}_k)_{m,m} \middle| \mathbf{g} = 1 \right]} \right) + \frac{r}{N \beta} \log(e^{-\eta} - e^{-\xi}). \quad (80)$$

where \mathbf{u} is a proper complex Gaussian random variable ($\mathbf{u} \sim \mathcal{CN}(0, 1)$) and η and ξ are the thresholds used to define the truncated Gaussian distribution.

As in the previous bounds, this lower bound depends on the conditional channel distribution (through the expectation of the conditional channel covariance matrix). Note that the covariance matrix itself also depends on the definition of the truncated Gaussian distribution, i.e., on the parameters η and ξ .

Proof of Lemma 3.a: See Appendix A.

VII. CHANNEL ESTIMATION

In this section we analyze the conditional distribution of the channel given the past transmitted and received symbols, and provide the bounds required for the capacity evaluation.

Using Assumption III (no feedback) the transmitted signal is statistically independent from the channel, and hence, given the input symbols $\tilde{\mathbf{X}}_0^{k-1}$, the output symbols $\tilde{\mathbf{Y}}_0^{k-1}$ are jointly Gaussian with the channel $\tilde{\mathbf{H}}_0^k$. Thus, the desired distribution can be analyzed through the theory of optimal linear estimation. The auto-covariance of the conditioned output symbols, $\tilde{\mathbf{Y}}_0^{k-1} | \tilde{\mathbf{X}}_0^{k-1}$, is:

$$\mathbf{C}_{\tilde{\mathbf{Y}}_0^{k-1} | \tilde{\mathbf{X}}_0^{k-1}} = N \text{diag}(\tilde{\mathbf{X}}_0^{k-1}) \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} \text{diag}(\tilde{\mathbf{X}}_0^{k-1})^\dagger + \mathbf{I} \quad (81)$$

and their cross covariance matrix with $\tilde{\mathbf{H}}_k$ is:

$$\mathbf{C}_{\tilde{\mathbf{Y}}_0^{k-1}, \tilde{\mathbf{H}}_k | \tilde{\mathbf{X}}_0^{k-1}} = \sqrt{N} \cdot \text{diag}(\tilde{\mathbf{X}}_0^{k-1}) \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}, \tilde{\mathbf{H}}_k} \quad (82)$$

where $\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}, \tilde{\mathbf{H}}_k}$ is the sub-matrix of $\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}}$ describing the cross covariance between current channel vector and the past channel vectors. Therefore, $\tilde{\mathbf{H}}_k | \tilde{\mathbf{X}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1}$ has complex Gaussian distribution $\tilde{\mathbf{H}}_k | \tilde{\mathbf{X}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1} \sim \mathcal{CN}(\tilde{\boldsymbol{\mu}}_k, \tilde{\boldsymbol{\Sigma}}_k)$, where the conditional channel mean, $\tilde{\boldsymbol{\mu}}_k$, and the conditional channel covariance matrix, $\tilde{\boldsymbol{\Sigma}}_k$, are given by [32]:

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_k &= E[\tilde{\mathbf{H}}_k] + \sqrt{N} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}, \tilde{\mathbf{H}}_k}^\dagger \text{diag}(\tilde{\mathbf{X}}_0^{k-1})^\dagger \left(N \text{diag}(\tilde{\mathbf{X}}_0^{k-1}) \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} \text{diag}(\tilde{\mathbf{X}}_0^{k-1})^\dagger + \mathbf{I} \right)^{-1} \\ &\quad \cdot \left(\tilde{\mathbf{Y}}_0^{k-1} - \sqrt{N} \text{diag}(\tilde{\mathbf{X}}_0^{k-1}) E[\tilde{\mathbf{H}}_0^{k-1}] \right) \end{aligned} \quad (83)$$

$$\begin{aligned} \tilde{\mathbf{E}}_k &= \mathbf{C}_{\tilde{\mathbf{H}}_k} - N \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}, \tilde{\mathbf{H}}_k}^\dagger \text{diag}(\tilde{\mathbf{X}}_0^{k-1})^\dagger \left(N \text{diag}(\tilde{\mathbf{X}}_0^{k-1}) \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} \text{diag}(\tilde{\mathbf{X}}_0^{k-1})^\dagger + \mathbf{I} \right)^{-1} \\ &\quad \cdot \text{diag}(\tilde{\mathbf{X}}_0^{k-1}) \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}, \tilde{\mathbf{H}}_k}. \end{aligned} \quad (84)$$

We note that the conditional channel mean is characterized by the conditional channel covariance matrix, $\tilde{\mathbf{E}}_k$, since we can write the distribution of the channel mean as $\tilde{\boldsymbol{\mu}}_k | \tilde{\mathbf{X}}_0^{k-1} \sim \mathcal{CN} \left(E[\tilde{\mathbf{H}}_k], \mathbf{C}_{\tilde{\mathbf{H}}_k} - \tilde{\mathbf{E}}_k \right)$. Therefore, we will focus hereon on the characterization of the conditional channel covariance matrix $\tilde{\mathbf{E}}_k$.

In this section we characterize the bounds on $(\tilde{\mathbf{E}}_k)_{m,m}$ and its moments in terms of the effective coherence time.

Lemma 1.b: With a peak power constraint, the conditional channel covariance matrix is lower bounded by:

$$\lim_{k \rightarrow \infty} \min_m \inf_{\tilde{\mathbf{X}}_0^{k-1} \in \mathcal{I}(N p_x \mathbf{I})} (\tilde{\mathbf{E}}_k)_{m,m} \geq \frac{1}{N} \frac{1}{1 + \frac{p_x}{2} (\hat{T}_c(p_x) - N)}. \quad (85)$$

With a quadratic power constraint, the conditional channel covariance matrix satisfies:

$$\lim_{k \rightarrow \infty} \sup_{\text{Pr}(\mathcal{P}_{k-1})} E \left[\left\{ 1 - N \min_m \inf_{\tilde{\mathbf{X}}_0^{k-1} \in \mathcal{I}(\mathcal{P}_{k-1})} (\tilde{\mathbf{E}}_k)_{m,m} \right\}^2 \right] \leq \alpha \left(1 - \frac{1}{1 + \frac{p_x}{2} (\hat{T}_c(p_x) - N)} \right)^2, \quad (86)$$

Proof of Lemma 1.b: In Appendix B we prove the following inequality:

$$(\tilde{\mathbf{E}}_k)_{m,m} \geq \frac{1}{N} \left[1 - D_k^\dagger \mathcal{P}_{k-1} (\mathbf{I} + \Delta_{k-1} \mathcal{P}_{k-1})^{-1} D_k \right]. \quad (87)$$

The proof is based on showing that the conditional covariance is minimized using $\mathcal{S}_{\mathbf{X}_0^{k-1}} = \mathcal{P}_{k-1} \otimes \mathbf{J}_m$, where \mathbf{J}_m is an $N \times N$ matrix with only one nonzero element, the m -th element on the diagonal, which equals 1.

The first part of the lemma follows directly from (87) by noting that the conditional covariance matrix is a nonincreasing function of each element in \mathcal{P}_{k-1} . Therefore, it is minimized when all elements take their maximal value ($\mathcal{P}_{k-1} = N p_x \mathbf{I}$), so that

$$\lim_{k \rightarrow \infty} \min_m \inf_{\tilde{\mathbf{X}}_0^{k-1} \in \mathcal{I}(N p_x \mathbf{I})} (\tilde{\mathbf{E}}_k)_{m,m} \geq \frac{1}{N} \left[1 - \lim_{k \rightarrow \infty} N p_x D_k^\dagger (\mathbf{I} + N p_x \Delta_{k-1})^{-1} D_k \right]. \quad (88)$$

Using the definition of the effective coherence time, (12), and the matrix inversion lemma leads to (85) and completes the first part of the proof.

For the quadratic case the bounding is slightly more complicated, and we search for bounds on moments of the conditional covariance matrix that will hold for any distribution of the

transmission powers, \mathcal{P}_{k-1} . To do so, we use the same derivation as in (B.4), and apply it to (87) with $\mathcal{P}_{k-1} = E[\mathcal{P}_{k-1}] + \check{\mathcal{P}}$. We get:

$$\begin{aligned} (\tilde{\mathbf{e}}_k)_{m,m} &\geq \frac{1}{N} - \frac{1}{N} D_k^\dagger E[\mathcal{P}_{k-1}] (\mathbf{I} + \Delta_{k-1} E[\mathcal{P}_{k-1}])^{-1} D_k \\ &\quad - \frac{1}{N} D_k^\dagger (\mathbf{I} + E[\mathcal{P}_{k-1}] \Delta_{k-1})^{-1} \check{\mathcal{P}} (\mathbf{I} + \Delta_{k-1} E[\mathcal{P}_{k-1}])^{-1} D_k. \end{aligned} \quad (89)$$

Taking the expectation of the square of (89) and using $(\tilde{\mathbf{e}}_k)_{m,m} \leq \frac{1}{N}$, we have:

$$E \left[\left\{ 1 - N(\tilde{\mathbf{e}}_k)_{m,m} \right\}^2 \right] \leq \left(D_k^\dagger E[\mathcal{P}_{k-1}] (\mathbf{I} + \Delta_{k-1} E[\mathcal{P}_{k-1}])^{-1} D_k \right)^2 + E \left[V^\dagger \check{\mathcal{P}} V V^\dagger \check{\mathcal{P}} V \right]. \quad (90)$$

where $V = (\mathbf{I} + \Delta_{k-1} E[\mathcal{P}_{k-1}])^{-1} D_k$. Focusing on the second term in the righthand side of (90) we have:

$$\begin{aligned} E \left[V^\dagger \check{\mathcal{P}} V V^\dagger \check{\mathcal{P}} V \right] &= E \left[\sum_a \sum_b v_a^* \check{\mathcal{P}}_{a,a} v_a v_b^* \check{\mathcal{P}}_{b,b} v_b \right] \\ &= \sum_a \sum_b |v_a|^2 |v_b|^2 E \left[\check{\mathcal{P}}_{a,a} \check{\mathcal{P}}_{b,b} \right] \\ &\leq \sum_a \sum_b |v_a|^2 |v_b|^2 \sqrt{E \left[\check{\mathcal{P}}_{a,a}^2 \right] E \left[\check{\mathcal{P}}_{b,b}^2 \right]} \\ &\leq (\alpha - 1) N^2 p_x^2 (V^\dagger V)^2 \end{aligned} \quad (91)$$

where we use $\check{\mathcal{P}}_{b,b} = \mathbf{p}_b - E[\mathbf{p}_b]$, and hence $E \left[\check{\mathcal{P}}_{b,b}^2 \right] \leq (\alpha - 1) N^2 p_x^2$. Substituting back in (90), we observe that the righthand side of (90) is maximized with $E[\mathcal{P}_{k-1}] = N p_x \mathbf{I}$ and we get:

$$\begin{aligned} E \left[\left\{ 1 - N(\tilde{\mathbf{e}}_k)_{m,m} \right\}^2 \right] &\leq N^2 p_x^2 \left(D_k^\dagger (\mathbf{I} + N p_x \Delta_{k-1})^{-1} D_k \right)^2 \\ &\quad + (\alpha - 1) N^2 p_x^2 \left(D_k^\dagger (\mathbf{I} + N p_x \Delta_{k-1})^{-2} D_k \right)^2 \\ &\leq N^2 p_x^2 \left(D_k^\dagger (\mathbf{I} + N p_x \Delta_{k-1})^{-1} D_k \right)^2 \\ &\quad + (\alpha - 1) N^2 p_x^2 \left(D_k^\dagger (\mathbf{I} + N p_x \Delta_{k-1})^{-1} D_k \right)^2 \\ &= \alpha N^2 p_x^2 \left(D_k^\dagger (\mathbf{I} + N p_x \Delta_{k-1})^{-1} D_k \right)^2 \end{aligned} \quad (92)$$

where the second inequality in (92) uses the fact that all eigenvalues of $\mathbf{I} + N p_x \Delta_{k-1}$ are larger or equal to 1, and therefore $(\mathbf{I} + N p_x \Delta_{k-1})^{-1} \geq (\mathbf{I} + N p_x \Delta_{k-1})^{-2}$. Using the effective coherence time definition results in (86) and completes the proof of the lemma. ■

Lemma 2.b: Using QPSK modulation, as defined in (66), for each active frequency bin ($0 \leq m \leq r-1$) the diagonal element of the conditional channel covariance matrix is lower bounded by:

$$\lim_{k \rightarrow \infty} (\tilde{\mathbf{E}}_k)_{m,m} \leq \begin{cases} \frac{1}{N} \frac{1}{1 + \frac{\beta p_x}{L} \frac{1}{2} (\hat{T}_c (\beta p_x) - N)} & r = N \\ \frac{1}{N} \frac{1}{1 + \frac{\beta p_x}{r} \frac{1}{2} (\hat{T}_c (\frac{\beta}{r} p_x) - N)} & r < N \end{cases} \quad (93)$$

Proof of Lemma 2.b: Starting with the wideband QPSK modulation ($r = N$), the transmission spectrum given $\mathbf{g} = 1$ is equal to $\mathcal{S}_{\mathbf{X}_0^{k-1}} = \beta p_x \mathbf{I}$. Substituting in the estimation error covariance matrix, (84) (see also (B.3)), we have:

$$\tilde{\mathbf{E}}_k = \mathbf{C}_{\tilde{\mathbf{H}}_k} - N\beta p_x \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}, \tilde{\mathbf{H}}_k}^\dagger \left(N\beta p_x \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \mathbf{I} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}, \tilde{\mathbf{H}}_k}. \quad (94)$$

Writing the estimation error covariance matrix, (94), in the time domain, and substituting (7) we have:

$$\mathbf{E}_k = \mathbf{F} \tilde{\mathbf{E}}_k \mathbf{F}^\dagger = \mathbf{C}_{\mathbf{H}_k} - N\beta p_x D_k^\dagger \otimes \mathbf{C}_{\mathbf{H}_k} \left(N\beta p_x \Delta_{k-1} \otimes \mathbf{C}_{\mathbf{H}_k} + \mathbf{I} \right)^{-1} D_k \otimes \mathbf{C}_{\mathbf{H}_k}. \quad (95)$$

Note that covariance matrices are hermitian so that $\mathbf{C}_{\mathbf{H}_k}^\dagger = \mathbf{C}_{\mathbf{H}_k}$. In Appendix C we show that this is a diagonal matrix in which the l -th element on the diagonal is given by:

$$(\mathbf{E}_k)_{l,l} = (\mathbf{C}_{\mathbf{H}_k})_{l,l} - N\beta p_x D_k^\dagger (\mathbf{C}_{\mathbf{H}_k})_{l,l} \left(N\beta p_x \Delta_{k-1} (\mathbf{C}_{\mathbf{H}_k})_{l,l} + \mathbf{I} \right)^{-1} D_k (\mathbf{C}_{\mathbf{H}_k})_{l,l}. \quad (96)$$

For any channel tap such that $(\mathbf{C}_{\mathbf{H}_k})_{l,l} > 0$ we can use the matrix inversion lemma to write the inverse of the estimation error as:

$$(\mathbf{E}_k)_{l,l}^{-1} = (\mathbf{C}_{\mathbf{H}_k})_{l,l}^{-1} + N\beta p_x D_k^\dagger \cdot \left(N\beta p_x (\mathbf{C}_{\mathbf{H}_k})_{l,l} \left[\Delta_{k-1} - D_k D_k^\dagger \right] + \mathbf{I} \right)^{-1} D_k \quad (97)$$

and taking the limit as k goes to infinity:

$$\begin{aligned} \lim_{k \rightarrow \infty} (\mathbf{E}_k)_{l,l}^{-1} &= (\mathbf{C}_{\mathbf{H}_k})_{l,l}^{-1} + \frac{\beta p_x}{2} \left(\hat{T}_c (\beta p_x (\mathbf{C}_{\mathbf{H}_k})_{l,l}) - N \right) \\ &\geq (\mathbf{C}_{\mathbf{H}_k})_{l,l}^{-1} + \frac{\beta p_x}{2} (\hat{T}_c (\beta p_x) - N). \end{aligned} \quad (98)$$

Going back to the frequency domain, the properties of the DFT guarantee that all elements on the diagonal of $\tilde{\mathbf{E}}_k$ are equal. We therefore have:

$$\lim_{k \rightarrow \infty} (\tilde{\mathbf{E}}_k)_{m,m} = \frac{1}{N} \text{Tr} \left\{ \lim_{k \rightarrow \infty} \mathbf{E}_k \right\} \leq \frac{1}{N} \sum_{l: (\mathbf{C}_{\mathbf{H}_k})_{l,l} \neq 0} \frac{1}{(\mathbf{C}_{\mathbf{H}_k})_{l,l}^{-1} + \beta \frac{p_x}{2} \left(\hat{T}_c (\beta p_x) - N \right)}. \quad (99)$$

Using the concavity of the function $(a + 1/x)^{-1}$ for positive a and $\text{Tr}[\mathbf{C}_{\mathbf{H}_k}] = 1$ proves the first (upper) part of (93).

For $r < N$ we lower bound the diagonal element of the conditional channel covariance matrix by the performance of a sub-optimal estimation scheme. This estimation scheme estimates the channel at the m -th frequency bin based only on data transmitted and received over the m frequency bin. Inspecting (94) and taking only the elements that correspond to the m frequency bin ($0 \leq m \leq r - 1$) the conditional channel variance is lower bounded by:

$$(\tilde{\mathbf{E}}_k)_{m,m} \leq \frac{1}{N} - \frac{1}{N} \beta \frac{N}{r} p_x D_k^\dagger \left(\beta \frac{N}{r} p_x \Delta_{k-1} + \mathbf{I} \right)^{-1} D_k. \quad (100)$$

where we took into account the normalization $(\mathbf{C}_{\tilde{\mathbf{H}}_k})_{m,m} = \frac{1}{N}$ and the fact that the transmitted power for an active bin, given $\mathbf{g} = 1$, is $\beta \frac{N}{r} p_x$. Using the matrix inversion lemma and the definition of the effective coherence time, (12), results in the second part of (93) and completes the proof of the lemma. ■

Lemma 3.b: Using truncated Gaussian input distribution, as defined in (78), for each active frequency bin ($0 \leq m \leq r - 1$) the mean of the diagonal element of the conditional channel covariance matrix is lower bounded for quadratic power constraint by:

$$\lim_{k \rightarrow \infty} E \left[(\tilde{\mathbf{E}}_k)_{m,m} \right] \leq \begin{cases} \frac{1}{N} \frac{1}{1 + \frac{\beta p_x}{(1+\eta) \log(1+\frac{1}{\eta})} \frac{1}{2L} \left(\hat{T}_c \left(\frac{\beta p_x}{(1+\eta) \log(1+\frac{1}{\eta})} \right) - N \right)} & r = N \\ \frac{1}{N} \frac{1}{1 + \frac{\beta p_x}{(1+\eta) \log(1+\frac{1}{\eta})} \frac{1}{2r} \left(\hat{T}_c \left(\frac{\beta p_x}{r(1+\eta) \log(1+\frac{1}{\eta})} \right) - N \right)} & r < N \end{cases} \quad (101)$$

and for peak power constraint by:

$$\lim_{k \rightarrow \infty} E \left[(\tilde{\mathbf{E}}_k)_{m,m} \right] \leq \begin{cases} \frac{1}{N} \frac{1}{1 + \frac{p_x}{\xi} \frac{1-e\eta-\xi}{\log(1+\frac{1}{\eta})} \frac{1}{2L} \left(\hat{T}_c \left(\frac{p_x}{\xi} \frac{1-e\eta-\xi}{\log(1+\frac{1}{\eta})} \right) - N \right)} & r = N \\ \frac{1}{N} \frac{1}{1 + \frac{p_x}{\xi} \frac{1-e\eta-\xi}{\log(1+\frac{1}{\eta})} \frac{1}{2r} \left(\hat{T}_c \left(\frac{p_x}{r\xi} \frac{1-e\eta-\xi}{\log(1+\frac{1}{\eta})} \right) - N \right)} & r < N \end{cases} \quad (102)$$

Proof of Lemma 3.b: Starting from the wideband truncated Gaussian distribution ($r = N$), the power transmitted in all frequency bins is strictly positive. We therefore can write the conditional channel covariance matrix (84) (see also (B.3)), as:

$$\tilde{\mathbf{E}}_k = f_1 \left(\frac{1}{N} \mathcal{S}_{\mathbf{X}_0^{k-1}}^{-1} \right) = \mathbf{C}_{\tilde{\mathbf{H}}_k} - \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k}^\dagger \left(\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \frac{1}{N} \mathcal{S}_{\mathbf{X}_0^{k-1}}^{-1} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k}. \quad (103)$$

Next we show that $f_1(\mathbf{A})$ is a concave function of each of the elements in the strictly positive diagonal matrix \mathbf{A} . Denote by \mathbf{I}_i the vector which is all zeros except for a 1 in the i -th element.

The derivative of $f_1(\mathbf{A})$ with respect to $(\mathbf{A})_{i,i}$ is given by:

$$\frac{\partial f_1(\mathbf{A})}{\partial (\mathbf{A})_{i,i}} = \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k}^\dagger \left(\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \mathbf{A} \right)^{-1} I_i I_i^\dagger \left(\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \mathbf{A} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k} \quad (104)$$

which is a positive semi-definite matrix. The second derivative is given by:

$$\frac{\partial^2 f_1(\mathbf{A})}{\partial (\mathbf{A})_{i,i}^2} = -2 \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k}^\dagger \left(\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \mathbf{A} \right)^{-1} I_i I_i^\dagger \left(\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \mathbf{A} \right)^{-1} I_i I_i^\dagger \left(\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \mathbf{A} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k} \quad (105)$$

which is a negative semi-definite matrix, and guaranties that the function $f_1(\mathbf{A})$ is a concave function of $(\mathbf{A})_{i,i}$.

In order to avoid the differentiation with respect to the whole matrix \mathbf{A} we considered only the derivatives with respect to each single element in it. But, as we showed that the function $f_1(\mathbf{A})$ is a concave function of each of the (diagonal) elements in \mathbf{A} regardless of the others elements, we can now apply the Jensen's inequality serially on one matrix element at a time. Going over all matrix elements we get:

$$E \left[\tilde{\boldsymbol{\epsilon}}_k \right] = E \left[f_1 \left(\frac{1}{N} \mathcal{S}_{\mathbf{X}_0^{k-1}}^{-1} \right) \right] \leq \mathbf{C}_{\tilde{\mathbf{H}}_k} - \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k}^\dagger \left(\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \frac{1}{N} E \left[\mathcal{S}_{\mathbf{X}_0^{k-1}}^{-1} \right] \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k}. \quad (106)$$

Taking into account the definition of the transmitted signal, (78), the expectation of the inverse of the transmitted spectrum, given $\mathbf{g} = 1$, is:

$$E \left[\mathcal{S}_{\mathbf{X}_0^{k-1}}^{-1} \right] = \mathbf{I} \cdot \frac{p_z}{c^2} \frac{1}{\theta} \int_{\eta}^{\xi} \chi^{-1} e^{-\chi} d\chi < \mathbf{I} \cdot \frac{p_z}{c^2} \frac{E_1(\eta)}{e^{-\eta} - e^{-\xi}} < \mathbf{I} \cdot \frac{p_z}{c^2} \frac{e^{-\eta} \log(1 + \frac{1}{\eta})}{e^{-\eta} - e^{-\xi}} \quad (107)$$

where $E_1()$ is the exponential integral function, and the second inequality used $E_1(a) < e^{-a} \log(1 + \frac{1}{a})$ [33].

Substituting (107) in (106) results in:

$$E \left[\tilde{\boldsymbol{\epsilon}}_k \right] \leq \mathbf{C}_{\tilde{\mathbf{H}}_k} - N \zeta \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k}^\dagger \left(N \zeta \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \mathbf{I} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1} \tilde{\mathbf{H}}_k} \quad (108)$$

where for quadratic power constraint, substituting (76), (A.21), $r = N$ and $\xi = \infty$ we have:

$$\zeta = \beta \frac{N}{r} p_x \frac{1}{(1 + \eta) \log(1 + \frac{1}{\eta})}. \quad (109)$$

For the peak power constraint, substituting (A.25) and $\beta = 1$

$$\zeta = \frac{N p_x}{r \xi} \frac{e^{-\eta} - e^{-\xi}}{e^{-\eta} \log(1 + \frac{1}{\eta})}. \quad (110)$$

Comparing (108) with (94) we note that the only difference is the change of the factor βp_x with the factor ζ . Hence, repeating the derivation of Lemma 2.b (Equations (95) to (99)) with the replaced factor results in:

$$\lim_{k \rightarrow \infty} E \left[(\tilde{\boldsymbol{\epsilon}}_k)_{m,m} \right] \leq \frac{1}{N} \frac{1}{1 + \frac{\zeta}{2L} (\hat{T}_c(\zeta) - N)}. \quad (111)$$

Substituting the proper values of ζ in (111) lead to the upper parts of (101) and (102) and completes the first part of the lemma.

Next for $r < N$. As in the QPSK case, we turn to the suboptimal single frequency bin estimation. In this case, the estimation error in the m -th frequency bin ($0 \leq m \leq r-1$) is given by:

$$\left(\tilde{\boldsymbol{\epsilon}}_k \right)_{m,m} \leq \frac{1}{N} - \frac{1}{N^2} D_k^\dagger \left(\frac{1}{N} \Delta_{k-1} + \frac{1}{N} \mathcal{A}_m \right)^{-1} D_k \quad (112)$$

where the diagonal matrix \mathcal{A}_m is generated by taking from the matrix $\mathcal{S}_{\mathbf{x}_0}^{-1}$ only the terms corresponding to the m -th element of each OFDM symbol.

Comparing (112) with (103), they have exactly the same structure, and hence we can repeat the first part of the proof (Equations (104) to (108)), adjusting only the corresponding matrices. The resulting bound is:

$$E \left[\tilde{\boldsymbol{\epsilon}}_k \right] \leq \frac{1}{N} - \frac{1}{N} \zeta D_k^\dagger (\zeta \Delta_{k-1} + \mathbf{I})^{-1} D_k. \quad (113)$$

with the values of ζ given by (109) and (110).

Using the matrix inversion lemma $(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$ and the definition of the effective coherence time, (12), results in the lower parts of (101) and (102) and completes the proof of the lemma. ■

VIII. CONCLUSIONS

In this paper we derived bounds on the capacity of the noncoherent stationary underspread complex Gaussian OFDM-WSSUS channel with a peak power constraint or a quadratic power constraint. The bounds are characterized only by the system signal-to-noise ratio (SNR) p_x and by a newly defined effective coherence time $\hat{T}_c(p_x)$, which measures the capability to estimate the channel and is a nonincreasing function of the system SNR. The bounds show that:

- The coherent channel capacity is achievable if $p_x \hat{T}_c(p_x) \gg 1$ and $\hat{T}_c(p_x) \gg N$.

- For the peak power constraint, if the channel has zero mean and $p_x \ll 1/\hat{T}_c(p_x)$ the capacity is approximately $\frac{1}{2}p_x^2\hat{T}_c(p_x)$.
- For the quadratic power constraint if the channel has zero mean and $p_x \ll 1/\hat{T}_c(p_x)$ the capacity ranges between $\frac{1}{2}p_x^2\hat{T}_c(p_x)$ and $\frac{\alpha}{2}p_x^2\hat{T}_c(p_x)$ (typically the effective coherence time changes slowly with SNR and the higher bound better characterizes the capacity). If the limit $\hat{T}_{c0} = \lim_{p_x \rightarrow 0} \hat{T}_c(p_x)$ exists, the capacity in the very low SNR limit is $\frac{\alpha}{2}p_x^2\hat{T}_c(p_x)$.
- As long as the effective coherence time is large enough ($\hat{T}_c(p_x) \gg N$) the channel capacity is achievable using a receiver that performs channel estimation based on past received and transmitted symbols, and decodes based on minimum Euclidean distance.

The paper presented an initial study of the effective coherence time and plotted it for the autoregressive AR1 channel. Owing to its relevance in the characterization of the channel capacity, it is important to better understand its properties. Future work is required to characterize the effective coherence time of commonly used channels, and to study the relationship between the effective coherence time and channel capacity in more complicated channel structures.

APPENDIX A

PROOF OF LEMMA 3.A

As stated in the proof of Lemma 2.a, we can safely assume that \mathbf{g} can be decoded with no error. We reuse the derivation of Lemma 2.a up to Equation (72), and focus on the mutual information of a single frequency bin:

$$C \geq \lim_{k \rightarrow \infty} \frac{1}{N\beta} \sum_{m=0}^{r-1} I(\tilde{\mathbf{z}}_{k,m}; \tilde{\mathbf{y}}_{k,m} | \tilde{\mathbf{Z}}_0^{k-1}, \tilde{\mathbf{Y}}_0^{k-1}, \mathbf{g} = 1).$$

In order to lower bound the mutual information we use the generalized mutual information (GMI, see for example [24]). The derivation of the GMI can give a lower bound on the mutual information evaluated by:

$$I(\mathbf{z}; \mathbf{y}) \geq E \left[\log \frac{e^{-sd(\mathbf{z}, \mathbf{y})}}{E[e^{-sd(\mathbf{z}', \mathbf{y})} | \mathbf{y}]} \right] \quad (\text{A.1})$$

where s is any positive constant, $d(\mathbf{z}, \mathbf{y})$ is any distance metric that defines the operation of the receiver (which chooses the codeword that minimizes the distance to the received signal), and \mathbf{z}' is a random variable with the same distribution of \mathbf{z} but statistically independent of \mathbf{z} and \mathbf{y} .

We next evaluate such lower bound based on the transmission of a truncated Gaussian distributed symbol over a flat fading Gaussian channel. Consider the channel:

$$\mathbf{y} = \mathbf{h}z + \mathbf{v} \quad (\text{A.2})$$

where $z \sim \mathcal{TCN}(\eta, \xi)$ is the input symbol with truncated Gaussian distribution, \mathbf{h} is the Gaussian channel with $\mathbf{h} \sim \mathcal{CN}(\mu, \epsilon)$ and \mathbf{v} is an additive white proper complex Gaussian noise with $\mathbf{v} \sim \mathcal{CN}(0, 1)$. We assume a conventional coherent receiver, i.e., the distance metric is:

$$d(z, y) = |y - \mu z|^2. \quad (\text{A.3})$$

The more difficult part in the evaluation of (A.1) is typically the evaluation of the expectation in the denominator. Luckily, we can use the fact that the input distribution is close to the Gaussian distribution. For a proper complex Gaussian random variable $\mathbf{u} \sim \mathcal{CN}(0, 1)$ we have:

$$E \left[e^{-|\nu - a\mathbf{u}|^2} \right] = \frac{1}{1 + |a|^2} e^{-\frac{|\nu|^2}{1 + |a|^2}}. \quad (\text{A.4})$$

As the expression inside the expectation is always positive, we can lower bound this expectation using the truncated Gaussian distribution by:

$$\frac{1}{1 + |a|^2} e^{-\frac{|\nu|^2}{1 + |a|^2}} \geq \theta E \left[e^{-|\nu - a\mathbf{u}|^2} \middle| \eta \leq |\mathbf{u}|^2 \leq \xi \right]. \quad (\text{A.5})$$

where $\theta = e^{-\eta} - e^{-\xi}$ is the probability that $\eta \leq |\mathbf{u}|^2 \leq \xi$. Substituting $\nu = \sqrt{s}\mathbf{y}$ and $a = \sqrt{s}\mu$, we have for a truncated Gaussian distributed \mathbf{z}' :

$$E \left[e^{-sd(\mathbf{z}', \mathbf{y})} \middle| \mathbf{y} \right] \leq \frac{1}{e^{-\eta} - e^{-\xi}} \frac{1}{1 + s|\mu|^2} e^{-\frac{s|\mathbf{y}|^2}{1 + s|\mu|^2}} \quad (\text{A.6})$$

Substituting (A.6) in (A.1), the resulting bound is:

$$\begin{aligned} I(\mathbf{z}; \mathbf{y}) &\geq \log(e^{-\eta} - e^{-\xi}) + \log(1 + s|\mu|^2) + E \left[-s|\mathbf{y} - \mu\mathbf{z}|^2 + \frac{s|\mathbf{y}|^2}{1 + s|\mu|^2} \right] \\ &= \log(e^{-\eta} - e^{-\xi}) + \log(1 + s|\mu|^2) - s(1 + p_z\epsilon) + \frac{s(|\mu|^2 p_z + 1 + p_z\epsilon)}{1 + s|\mu|^2}. \end{aligned} \quad (\text{A.7})$$

We arbitrarily choose s to zero the last 2 terms in (A.7), i.e.:

$$s = \frac{p_z}{1 + p_z\epsilon}, \quad (\text{A.8})$$

which results in:

$$I(\mathbf{z}; \mathbf{y}) \geq \log(e^{-\eta} - e^{-\xi}) + \log \left(1 + \frac{p_z|\mu|^2}{1 + p_z\epsilon} \right). \quad (\text{A.9})$$

To use this bound for the actual channel at hand and the modulation described by (78), we substitute $\mu = c\tilde{\mu}_{k,m}\sqrt{N/p_z}$ and $\epsilon = Nc^2(\tilde{\epsilon}_k)_{m,m}/p_z$, which results in the bound:

$$C \geq \lim_{k \rightarrow \infty} \frac{1}{N\beta} \sum_{m=0}^{r-1} E \left[\log \left(1 + \frac{Nc^2 |\tilde{\mu}_{k,m}|^2}{1 + Nc^2 (\tilde{\epsilon}_k)_{m,m}} \right) \middle| \mathbf{g} = 1 \right] + \frac{r}{N\beta} \log(e^{-\eta} - e^{-\xi}). \quad (\text{A.10})$$

Next, we show that (A.10) is convex with respect to $(\tilde{\epsilon}_k)_{m,m}$, and hence can be lower bounded using the expectation of $(\tilde{\epsilon}_k)_{m,m}$. To do so we use the distribution of $\tilde{\mu}_{k,m}$, (49). At this point we also use $E[\tilde{\mathbf{h}}_{k,m}] = 0$, as was stated in the body of the lemma. We introduce a new proper complex Gaussian random variable, $\mathbf{u} \sim \mathcal{CN}(0, 1)$, statistically independent of $(\tilde{\epsilon}_k)_{m,m}$, and note that the distribution of $\tilde{\mu}_{k,m}$ is identical to the distribution of $\mathbf{u}\sqrt{\frac{1}{N} - (\tilde{\epsilon}_k)_{m,m}}$. Using the random variable \mathbf{u} we can rewrite (A.10) as:

$$\begin{aligned} C &\geq \lim_{k \rightarrow \infty} \frac{1}{N\beta} \sum_{m=0}^{r-1} E \left[\log \left(1 + \frac{Nc^2 \left(\frac{1}{N} - (\tilde{\epsilon}_k)_{m,m} \right) |\mathbf{u}|^2}{1 + Nc^2 (\tilde{\epsilon}_k)_{m,m}} \right) \middle| \mathbf{g} = 1 \right] + \frac{r}{N\beta} \log(e^{-\eta} - e^{-\xi}) \\ &= \frac{r}{N\beta} \log(e^{-\eta} - e^{-\xi}) + \lim_{k \rightarrow \infty} \frac{1}{N\beta} \sum_{m=0}^{r-1} \\ &\quad E \left[E \left[\log \left(1 + \frac{Nc^2 \left(\frac{1}{N} - (\tilde{\epsilon}_k)_{m,m} \right) |\mathbf{u}|^2}{1 + Nc^2 (\tilde{\epsilon}_k)_{m,m}} \right) \middle| (\tilde{\epsilon}_k)_{m,m} \right] \middle| \mathbf{g} = 1 \right]. \end{aligned} \quad (\text{A.11})$$

where in the second equation we used the law of total expectations, and the inner expectation is taken only with respect to \mathbf{u} .

We define the function:

$$f(\epsilon) = E \left[\log \left(1 + \frac{Nc^2 \left(\frac{1}{N} - \epsilon \right) |\mathbf{u}|^2}{1 + Nc^2 \epsilon} \right) \right] \quad (\text{A.12})$$

and rewrite (A.11) as:

$$C \geq \lim_{k \rightarrow \infty} \frac{1}{N\beta} \sum_{m=0}^{r-1} E \left[f \left((\tilde{\epsilon}_k)_{m,m} \right) \middle| \mathbf{g} = 1 \right] + \frac{r}{N\beta} \log(e^{-\eta} - e^{-\xi}). \quad (\text{A.13})$$

We next show that $f(\epsilon)$ is convex for $0 \leq \epsilon \leq \frac{1}{N}$ by proving that its second derivative with respect to ϵ is nonnegative. The first derivative is given by:

$$\begin{aligned} \frac{df(\epsilon)}{d\epsilon} &= E \left[\frac{-Nc^2 |\mathbf{u}|^2 (1 + Nc^2 \epsilon) - Nc^2 \left(Nc^2 \left(\frac{1}{N} - \epsilon \right) |\mathbf{u}|^2 \right)}{(1 + Nc^2 \epsilon + Nc^2 \left(\frac{1}{N} - \epsilon \right) |\mathbf{u}|^2) (1 + Nc^2 \epsilon)} \right] \\ &= Nc^2 E \left[\frac{-(1 + c^2) |\mathbf{u}|^2}{(1 + Nc^2 \epsilon + Nc^2 \left(\frac{1}{N} - \epsilon \right) |\mathbf{u}|^2) (1 + Nc^2 \epsilon)} \right]. \end{aligned} \quad (\text{A.14})$$

The second derivative is:

$$\begin{aligned} \frac{d^2 f(\epsilon)}{d\epsilon^2} = N^2 c^4 E & \left[\frac{(1+c^2)|\mathbf{u}|^2 (1-|\mathbf{u}|^2) (1+Nc^2\epsilon)}{(1+Nc^2\epsilon + Nc^2 (\frac{1}{N} - \epsilon) |\mathbf{u}|^2)^2 (1+Nc^2\epsilon)^2} \right] \\ & + N^2 c^4 E \left[\frac{(1+c^2)|\mathbf{u}|^2 (1+Nc^2\epsilon + Nc^2 (\frac{1}{N} - \epsilon) |\mathbf{u}|^2)}{(1+Nc^2\epsilon + Nc^2 (\frac{1}{N} - \epsilon) |\mathbf{u}|^2)^2 (1+Nc^2\epsilon)^2} \right] \end{aligned} \quad (\text{A.15})$$

Inspecting the expectation in the second line of (A.15), the term inside the expectation is always positive. Using $\epsilon < \frac{1}{N}$ we can lower bound this expectation by removing the term $Nc^2 (\frac{1}{N} - \epsilon) |\mathbf{u}|^2$ from the numerator. The resulting expectation is quite similar to the expectation in the first line of (A.15). Combining the two expectation results in:

$$\frac{d^2 f(\epsilon)}{d\epsilon^2} \geq N^2 c^4 E \left[\frac{(1+c^2)|\mathbf{u}|^2 (2-|\mathbf{u}|^2)}{(1+Nc^2\epsilon + Nc^2 (\frac{1}{N} - \epsilon) |\mathbf{u}|^2)^2 (1+Nc^2\epsilon)} \right]. \quad (\text{A.16})$$

We next use the inequality $(a-x)/(b+x) \geq (a-x)/(b+a)$ which holds for $a, x > -b$. Slightly rewriting this inequality we have $(2-x)/(c+dx) \geq (2-x)/(c+2d)$, and using it in (A.16) we have:

$$\frac{d^2 f(\epsilon)}{d\epsilon^2} \geq \frac{N^2 c^4 (1+c^2)}{(1+Nc^2\epsilon + Nc^2 (\frac{1}{N} - \epsilon) \cdot 2)^2 (1+\beta \frac{N^2}{r} p_x \epsilon)} E [|\mathbf{u}|^2 (2-|\mathbf{u}|^2)]. \quad (\text{A.17})$$

To complete this part we note that the multiplicative term in (A.17) is positive, and the expectation is easily evaluated using $E[|\mathbf{u}|^2] = 1$ and $E[|\mathbf{u}|^4] = 2$. Thus, the second derivative is always nonnegative:

$$\frac{d^2 f(\epsilon)}{d\epsilon^2} \geq 0. \quad (\text{A.18})$$

Equation (A.18) shows that the function $f(\epsilon)$ is convex. Using the Jensen's inequality we have:

$$E \left[f \left((\tilde{\mathbf{e}}_k)_{m,m} \right) \middle| \mathbf{g} = 1 \right] \geq f \left(E \left[(\tilde{\mathbf{e}}_k)_{m,m} \middle| \mathbf{g} = 1 \right] \right). \quad (\text{A.19})$$

Substituting (A.19) and (A.12) in (A.13) we get:

$$C \geq \lim_{k \rightarrow \infty} \frac{1}{N\beta} \sum_{m=0}^{r-1} \log \left(1 + \frac{Nc^2 \left(\frac{1}{N} - E \left[(\tilde{\mathbf{e}}_k)_{m,m} \middle| \mathbf{g} = 1 \right] \right) |\mathbf{u}|^2}{1 + Nc^2 E \left[(\tilde{\mathbf{e}}_k)_{m,m} \middle| \mathbf{g} = 1 \right]} \right) + \frac{r}{N\beta} \log(e^{-\eta} - e^{-\xi}). \quad (\text{A.20})$$

For the last stage of the proof we need to consider the different power constraints. For the quadratic constraint we set $\xi = \infty$ and c in (78) to:

$$c = \sqrt{\beta \frac{N}{r} p_x} \quad (\text{A.21})$$

using (78), (76) and (77) we have:

$$E[\tilde{\mathbf{X}}_k^\dagger \tilde{\mathbf{X}}_k] = Np_x \quad (\text{A.22})$$

$$E[(\tilde{\mathbf{X}}_k^\dagger \tilde{\mathbf{X}}_k)^2] = \frac{\beta N^2 p_x^2}{r^2} \frac{r(r-1)(1+\eta)^2 + r(2+\eta^2+2\eta)}{(1+\eta)^2} = \beta N^2 p_x^2 \left[1 + \frac{1}{r(1+\eta)^2} \right]. \quad (\text{A.23})$$

Adding to (A.20) the requirement to satisfy the quadratic power constraint results in (79) and proves the first part of the lemma.

For the peak power constraint we need to satisfy:

$$\frac{c^2}{p_z} r \xi \leq Np_x \quad (\text{A.24})$$

We thus set $\beta = 1$ and

$$c^2 = \frac{Np_x p_z}{r \xi}, \quad (\text{A.25})$$

which leads to (80) and completes the proof of the lemma.

APPENDIX B

MINIMIZATION OF THE CONDITIONAL COVARIANCE: PROOF OF EQUATION (87)

We search for the minimal estimation error in the m -th frequency bin given the transmitted power in each symbol. We use the following inequality which holds for the square matrices D , E and F if F and $D + E$ are positive semi-definite:

$$(D + E) (F(D + E) + I)^{-1} \leq D (FD + I)^{-1} + (DF + I)^{-1} E (FD + I)^{-1}. \quad (\text{B.1})$$

where $A \geq B$ means $A - B$ is positive semi-definite. To prove the inequality we write:

$$\begin{aligned} (D + E) (F(D + E) + I)^{-1} &= (D + E) (FD + I)^{-1} \\ &\quad - (D + E) (F(D + E) + I)^{-1} F E (FD + I)^{-1} \\ &= D (FD + I)^{-1} \\ &\quad + [I - (D + E) F ((D + E) F + I)^{-1}] E (FD + I)^{-1} \\ &= D (FD + I)^{-1} \\ &\quad + ((D + E) F + I)^{-1} E (FD + I)^{-1} \\ &= D (FD + I)^{-1} \\ &\quad + (DF + I)^{-1} E (FD + I)^{-1} \\ &\quad - (DF + I)^{-1} E F ((D + E) F + I)^{-1} E (FD + I)^{-1}, \quad (\text{B.2}) \end{aligned}$$

where the first and fourth equalities use the identity $(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - (\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}\mathbf{A}^{-1}$ and its transpose respectively, which hold for any nonsingular matrix \mathbf{A} , as long as $(\mathbf{A} + \mathbf{B})^{-1}$ exists. The second equality uses the identity $(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{I} + \mathbf{A}\mathbf{B})^{-1}$, which is obvious if the matrix \mathbf{B} is nonsingular, but holds for any positive semi-definite matrix \mathbf{B} as long as the inverse of $(\mathbf{I} + \mathbf{B}\mathbf{A})$ exists. The inequality in (B.1) results from the fact that the matrix $\mathbf{F}((\mathbf{D} + \mathbf{E})\mathbf{F} + \mathbf{I})^{-1}$ is positive semi-definite.

We use the identity $\mathbf{B}^\dagger(\mathbf{I} + \mathbf{B}\mathbf{A}\mathbf{B}^\dagger)^{-1}\mathbf{B} = \mathbf{B}^\dagger\mathbf{B}(\mathbf{I} + \mathbf{A}\mathbf{B}^\dagger\mathbf{B})^{-1}$, which holds as long as the inverse exists, and rewrite the estimation error covariance (84) as:

$$\tilde{\boldsymbol{\epsilon}}_k = \mathbf{C}_{\tilde{\mathbf{H}}_k} - N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k}^\dagger \mathcal{S}_{\mathbf{X}_0^{k-1}} \left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} \mathcal{S}_{\mathbf{X}_0^{k-1}} + \mathbf{I} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k}. \quad (\text{B.3})$$

Setting $\mathcal{S}_{\mathbf{X}_0^{k-1}} = \mathcal{J}_m + \mathcal{J}_{\bar{m}}$ where \mathcal{J}_m and $\mathcal{J}_{\bar{m}}$ are both diagonal, and using Inequality (B.1) results in:

$$\begin{aligned} \tilde{\boldsymbol{\epsilon}}_k &= \mathbf{C}_{\tilde{\mathbf{H}}_k} - N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k}^\dagger (\mathcal{J}_m + \mathcal{J}_{\bar{m}}) \left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} (\mathcal{J}_m + \mathcal{J}_{\bar{m}}) + \mathbf{I} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k} \\ &\geq \mathbf{C}_{\tilde{\mathbf{H}}_k} - N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k}^\dagger \mathcal{J}_m \left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} \mathcal{J}_m + \mathbf{I} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k} \\ &\quad - N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k}^\dagger \left(N\mathcal{J}_m \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} + \mathbf{I} \right)^{-1} \mathcal{J}_{\bar{m}} \left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} \mathcal{J}_m + \mathbf{I} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k} \end{aligned} \quad (\text{B.4})$$

Without loss of generality we will assume in this appendix that $m = 0$. Now assume that $\mathcal{J}_0 = \mathcal{P}_{k-1} \otimes \mathbf{J}_0$ and $\mathbf{J}_0 = \text{diag}([1, 0, \dots, 0])$. Also recall that $\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} = \Delta_{k-1} \otimes \mathbf{C}_{\tilde{\mathbf{H}}_0}$. Using the identity $(\mathbf{A} \otimes \mathbf{B}) \cdot (\mathbf{C} \otimes \mathbf{D}) = \mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D}$, the matrix inverse in the second line of (B.4) can be written as:

$$\left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} \mathcal{J}_0 + \mathbf{I} \right)^{-1} = \left(N(\Delta_{k-1} \mathcal{P}_{k-1}) \otimes (\mathbf{C}_{\tilde{\mathbf{H}}_0} \mathbf{J}_0) + \mathbf{I} \right)^{-1}. \quad (\text{B.5})$$

Noting that $(\mathbf{C}_{\tilde{\mathbf{H}}_0} \mathbf{J}_0)^2 = (\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0}(\mathbf{C}_{\tilde{\mathbf{H}}_0} \mathbf{J}_0)$, we use the identity:

$$(\mathbf{I} + \mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{I} - ((\mathbf{I} + \phi\mathbf{A})^{-1} \mathbf{A}) \otimes \mathbf{B} \quad (\text{B.6})$$

which holds if the inverse exists and $\mathbf{B}^2 = \phi\mathbf{B}$, and can be easily verified by direct multiplication. Using Identity (B.6), the inverse in (B.5) can be written as:

$$\left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}} \mathcal{J}_0 + \mathbf{I} \right)^{-1} = \mathbf{I} - \left\{ N \left(\mathbf{I} + N\Delta_{k-1} \mathcal{P}_{k-1} (\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0} \right)^{-1} \Delta_{k-1} \mathcal{P}_{k-1} \right\} \otimes (\mathbf{C}_{\tilde{\mathbf{H}}_0} \mathbf{J}_0). \quad (\text{B.7})$$

Substituting also $\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k} = D_k \otimes \mathbf{C}_{\tilde{\mathbf{H}}_0}$, we can write the multiplication:

$$\begin{aligned} & \left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k} \mathcal{J}_0 + \mathbf{I} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k} \\ &= D_k \otimes \mathbf{C}_{\tilde{\mathbf{H}}_0} - N \left\{ \left(\mathbf{I} + N\Delta_{k-1}\mathcal{P}_{k-1}(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0} \right)^{-1} \Delta_{k-1}\mathcal{P}_{k-1}D_k \right\} \otimes \left(\mathbf{C}_{\tilde{\mathbf{H}}_0} \mathbf{J}_0 \mathbf{C}_{\tilde{\mathbf{H}}_0} \right). \end{aligned} \quad (\text{B.8})$$

As we test the estimation error in the first frequency bin, we need to evaluate the first column of the matrix $\left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k} \mathcal{J}_0 + \mathbf{I} \right)^{-1} \mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k}$. This column is given by:

$$\begin{aligned} & v_{aN+b} \\ &= (D_k)_a (\mathbf{C}_{\tilde{\mathbf{H}}_0})_{b,0} - N \left(\left(\mathbf{I} + N\Delta_{k-1}\mathcal{P}_{k-1}(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0} \right)^{-1} \Delta_{k-1}\mathcal{P}_{k-1}D_k \right)_a (\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0} (\mathbf{C}_{\tilde{\mathbf{H}}_0})_{b,0} \\ &= g(a) (\mathbf{C}_{\tilde{\mathbf{H}}_0})_{b,0} \end{aligned} \quad (\text{B.9})$$

Now the top left element of the matrix described by the last line of (B.4) can be written as:

$$\begin{aligned} & -N \sum_{a=0}^{k-1} \sum_{b=0}^{N-1} |v_{aN+b}|^2 (\mathcal{J}_{\bar{0}})_{aN+b, aN+b} \\ &= -N \sum_{a=0}^{k-1} |g(a)|^2 \sum_{b=0}^{N-1} |(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{b,0}|^2 (\mathcal{J}_{\bar{0}})_{aN+b, aN+b} \\ &= -N \sum_{a=0}^{k-1} |g(a)|^2 \left[|(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0}|^2 (\mathcal{J}_{\bar{0}})_{aN, aN} + \sum_{b=1}^{N-1} |(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{b,0}|^2 (\mathcal{J}_{\bar{0}})_{aN+b, aN+b} \right] \\ &\geq 0 \end{aligned} \quad (\text{B.10})$$

where we use the fact that $\mathcal{S}_{\mathbf{X}_0^{k-1}} = \mathcal{J}_0 + \mathcal{J}_{\bar{0}}$ and $\text{Tr}(\mathcal{S}_{\mathbf{X}_a}) \leq \mathbf{p}_a$, and therefore $(\mathcal{J}_{\bar{0}})_{aN, aN} \leq 0$ and $\sum_{b=0}^{N-1} (\mathcal{J}_{\bar{0}})_{aN+b, aN+b} = 0$. We also use the positive semi-definiteness of the matrix $\mathbf{C}_{\tilde{\mathbf{H}}_0}$ so that $|(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0}|^2 \geq |(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{b,0}|^2$.

We conclude that the bound on the estimation error for the first frequency bin is achieved when $\mathcal{J}_{\bar{0}} = 0$. Using $\mathbf{J}_0 \mathbf{C}_{\tilde{\mathbf{H}}_0} \mathbf{J}_0 = (\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0} \mathbf{J}_0$ and (B.7) we evaluate the matrix:

$$\begin{aligned} & \mathcal{J}_0 \cdot \left(N\mathbf{C}_{\tilde{\mathbf{H}}_0^{k-1}\tilde{\mathbf{H}}_k} \mathcal{J}_0 + \mathbf{I} \right)^{-1} = \mathcal{P}_{k-1} \otimes \mathbf{J}_0 \\ & \quad - \left\{ \mathcal{P}_{k-1} N \left(\mathbf{I} + N\Delta_{k-1}\mathcal{P}_{k-1}(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0} \right)^{-1} \Delta_{k-1}\mathcal{P}_{k-1}(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0} \right\} \otimes \mathbf{J}_0 \\ &= \left\{ \mathcal{P}_{k-1} \left(\mathbf{I} + N\Delta_{k-1}\mathcal{P}_{k-1}(\mathbf{C}_{\tilde{\mathbf{H}}_0})_{0,0} \right)^{-1} \right\} \otimes \mathbf{J}_0. \end{aligned} \quad (\text{B.11})$$

Substituting (B.11) and $\mathcal{J}_0 = 0$ in (B.3), the bound is given by:

$$\begin{aligned}
(\tilde{\mathcal{E}}_k)_{0,0} &\geq (C_{\tilde{\mathbf{H}}_0})_{0,0} - N \left(\left(D_k^\dagger \otimes C_{\tilde{\mathbf{H}}_0} \right) \mathcal{J}_0 \left(N C_{\tilde{\mathbf{H}}_0}^{k-1} \mathcal{S}_{\mathbf{X}_0^{k-1}} + \mathbf{I} \right)^{-1} \left(D_k \otimes C_{\tilde{\mathbf{H}}_0} \right) \right)_{0,0} \\
&= (C_{\tilde{\mathbf{H}}_0})_{0,0} - N D_k^\dagger \mathcal{P}_{k-1} \left(\mathbf{I} + N \Delta_{k-1} \mathcal{P}_{k-1} (C_{\tilde{\mathbf{H}}_0})_{0,0} \right)^{-1} D_k \cdot \left(C_{\tilde{\mathbf{H}}_0} \mathbf{J}_0 C_{\tilde{\mathbf{H}}_0} \right)_{0,0} \\
&= (C_{\tilde{\mathbf{H}}_0})_{0,0} - N (C_{\tilde{\mathbf{H}}_0})_{0,0}^2 D_k^\dagger \mathcal{P}_{k-1} \left(\mathbf{I} + N \Delta_{k-1} \mathcal{P}_{k-1} (C_{\tilde{\mathbf{H}}_0})_{0,0} \right)^{-1} D_k. \tag{B.12}
\end{aligned}$$

Using our normalization $(C_{\tilde{\mathbf{H}}_0})_{0,0} = 1/N$, and we get (87).

APPENDIX C

DERIVATION OF EQUATION (96)

In this appendix we present the derivation of Equation (96) from Equation (95). The easiest way to do so is to define a matrix modulo concatenation, and use its properties to simplify the derivation. Note that the matrix modulo concatenation can be seen as a permutation of a block diagonal matrix concatenation. Therefore the two concatenations share the same properties. Yet, for the problems at hand, the modulo concatenation is more convenient as the matrices we handle have the desired form.

For the matrices of identical size $\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{v-1}$ we define the modulo concatenation as $\mathbf{A} = \backslash \mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{v-1} \backslash$ so that its elements satisfy:

$$(\mathbf{A})_{k,l} = \begin{cases} 0 & k \bmod v \neq l \bmod v \\ (\mathbf{A}_{k \bmod v})_{\lfloor \frac{k}{v} \rfloor, \lfloor \frac{l}{v} \rfloor} & k \bmod v = l \bmod v \end{cases}, \tag{C.1}$$

i.e., the elements of the concatenated matrices are placed on the diagonals of the blocks of the matrix \mathbf{A} . If the size of the matrices is $r \times N$ then the size of their diagonal concatenation is $vr \times vN$.

The modulo concatenation has the following useful properties:

- 1) If $\mathbf{A} = \backslash \mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{v-1} \backslash$ and $\mathbf{B} = \backslash \mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{v-1} \backslash$ are the modulo concatenation of $r \times N$ matrices, then $\mathbf{A} + \mathbf{B} = \backslash \mathbf{A}_0 + \mathbf{B}_0, \mathbf{A}_1 + \mathbf{B}_1, \dots, \mathbf{A}_{v-1} + \mathbf{B}_{v-1} \backslash$.
- 2) If $\mathbf{A} = \backslash \mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_{v-1} \backslash$ is the modulo concatenation of $r \times N$ matrices, and $\mathbf{B} = \backslash \mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{v-1} \backslash$ is the modulo concatenation of $N \times U$ matrices, then $\mathbf{AB} = \backslash \mathbf{A}_0 \mathbf{B}_0, \mathbf{A}_1 \mathbf{B}_1, \dots, \mathbf{A}_{v-1} \mathbf{B}_{v-1} \backslash$.
- 3) $\mathbf{I} = \backslash \mathbf{I}, \mathbf{I}, \dots, \mathbf{I} \backslash$.
- 4) If the inverse exists then $\mathbf{A}^{-1} = \backslash \mathbf{A}_0^{-1}, \mathbf{A}_1^{-1}, \dots, \mathbf{A}_{v-1}^{-1} \backslash$.

5) If $D = \text{diag}([d_0, d_1, \dots, d_{v-1}])$ is a diagonal matrix, then $\mathbf{A} \otimes D = \backslash \mathbf{A}d_0, \mathbf{A}d_1, \dots, \mathbf{A}d_{v-1} \backslash$. Note that properties 1,3 and 5 follow directly from the definition of modulo concatenation, and property 4 follows from 2 and 3. Property 2 can be proved by permutating the matrices into block diagonal matrices. But, for the completeness of the treatment we give here a short direct proof:

The k, l element of the multiplication in property 2 is:

$$\begin{aligned} (\mathbf{AB})_{k,l} &= \sum_{w=0}^{Nv-1} \mathbf{A}_{k,w} \mathbf{B}_{w,l} \\ &= \sum_{w=0}^{N-1} \mathbf{A}_{k, wv + (k \bmod v)} \mathbf{B}_{wv + (k \bmod v), l} \end{aligned} \quad (\text{C.2})$$

which is 0 if $k \bmod v \neq l \bmod v$, if $k \bmod v = l \bmod v$ the k, l element is:

$$(\mathbf{AB})_{k,l} = \sum_{w=0}^{v-1} (\mathbf{A}_{k \bmod v})_{\lfloor \frac{k}{v} \rfloor, w} (\mathbf{B}_{k \bmod v})_{w, \lfloor \frac{l}{v} \rfloor} = (\mathbf{A}_{k \bmod v} \mathbf{B}_{k \bmod v})_{\lfloor \frac{k}{v} \rfloor, \lfloor \frac{l}{v} \rfloor} \quad (\text{C.3})$$

which completes the proof.

Using the properties of the modulo concatenation and noting that the matrix $\mathbf{C}_{\mathbf{H}_k}$ is diagonal, we can now rewrite Equation (95) as:

$$\begin{aligned} \hat{\mathcal{E}}_k &= \mathbf{C}_{\mathbf{H}_k} - N\beta p_x D_k^\dagger \otimes \mathbf{C}_{\mathbf{H}_k} (N\beta p_x \Delta_{k-1} \otimes \mathbf{C}_{\mathbf{H}_k} + \mathbf{I})^{-1} D_k \otimes \mathbf{C}_{\mathbf{H}_k} \\ &= \mathbf{C}_{\mathbf{H}_k} - N\beta p_x \backslash D_k^\dagger(\mathbf{C}_{\mathbf{H}_k})_{0,0}, \dots, D_k^\dagger(\mathbf{C}_{\mathbf{H}_k})_{N-1,N-1} \backslash \\ &\quad \cdot \backslash N\beta p_x \Delta_{k-1}(\mathbf{C}_{\mathbf{H}_k})_{0,0} + \mathbf{I}, \dots, N\beta p_x \Delta_{k-1}(\mathbf{C}_{\mathbf{H}_k})_{N-1,N-1} + \mathbf{I} \backslash^{-1} \\ &\quad \cdot \backslash D_k(\mathbf{C}_{\mathbf{H}_k})_{0,0}, \dots, D_k(\mathbf{C}_{\mathbf{H}_k})_{N-1,N-1} \backslash \\ &= \mathbf{C}_{\mathbf{H}_k} - N\beta p_x \backslash D_k^\dagger(\mathbf{C}_{\mathbf{H}_k})_{0,0} (N\beta p_x \Delta_{k-1}(\mathbf{C}_{\mathbf{H}_k})_{0,0} + \mathbf{I})^{-1} D_k(\mathbf{C}_{\mathbf{H}_k})_{0,0}, \dots \\ &\quad , D_k^\dagger(\mathbf{C}_{\mathbf{H}_k})_{N-1,N-1} (N\beta p_x \Delta_{k-1}(\mathbf{C}_{\mathbf{H}_k})_{N-1,N-1} + \mathbf{I})^{-1} D_k(\mathbf{C}_{\mathbf{H}_k})_{N-1,N-1} \backslash. \end{aligned} \quad (\text{C.4})$$

which leads to (96).

APPENDIX D

ALTERNATIVE TRUNCATED GAUSSIAN LOWER BOUND

In this appendix we derive an alternative bound that can replace the one in Theorem 3 without the need for the zero mean assumption. This bound is given by:

The capacity of a channel with a quadratic power constraint is lower bounded by:

$$C \geq \sup_{\eta > 0} \max_{r \leq N} \max_{1 \leq \beta \leq \frac{\alpha r(1+\eta)^2}{1+r(1+\eta)^2}} \text{LB}_{\text{TG}}^{(\text{qd})}(p_x, r, \beta, \eta) - \frac{r}{N} \eta \quad (\text{D.5})$$

and the capacity of a channel with a peak power constraint is lower bounded by:

$$C \geq \sup_{\eta > 0} \sup_{\xi > \eta} \max_{r \leq N} \text{LB}_{\text{TG}}^{(\text{pk})}(p_x, r, \eta, \xi) - \frac{r}{N} \log(e^{-\eta} - e^{-\xi}) \quad (\text{D.6})$$

where

$$\text{LB}_{\text{TG}}^{(\text{qd})}(p_x, r, \beta, \eta) = \begin{cases} \frac{r}{N\beta} \log \left(1 + \frac{\beta \frac{N^2}{r} p_x \left| E[\tilde{\mathbf{h}}_{0,0}] + \frac{\mathbf{u}}{\sqrt{N}} \sqrt{1 - \frac{1}{1 + \frac{\beta p_x \eta}{1+\eta} \frac{1}{2r} \left(\hat{T}_c \left(\frac{\beta p_x \eta}{r(1+\eta)} \right) - N \right)} \right|^2}{1 + \frac{\beta \frac{N}{r} p_x}{1 + \frac{\beta p_x \eta}{1+\eta} \frac{1}{2r} \left(\hat{T}_c \left(\frac{\beta p_x \eta}{r(1+\eta)} \right) - N \right)}} \right) & r < N \\ \frac{1}{\beta} \log \left(1 + \frac{\beta N p_x \left| E[\tilde{\mathbf{h}}_{0,0}] + \frac{\mathbf{u}}{\sqrt{N}} \sqrt{1 - \frac{1}{1 + \frac{\beta p_x \eta}{1+\eta} \frac{1}{2L} \left(\hat{T}_c \left(\frac{\beta p_x \eta}{1+\eta} \right) - N \right)} \right|^2}{1 + \frac{\beta p_x}{1 + \frac{\beta p_x \eta}{1+\eta} \frac{1}{2L} \left(\hat{T}_c \left(\frac{\beta p_x \eta}{1+\eta} \right) - N \right)}} \right) & r = N \end{cases}, \quad (\text{D.7})$$

$$\text{LB}_{\text{TG}}^{(\text{pk})}(p_x, r, \eta, \xi) = \begin{cases} \frac{r}{N} \log \left(1 + \frac{\frac{N^2 p_x p_z}{r \xi} \left| E[\tilde{\mathbf{h}}_{0,0}] + \frac{\mathbf{u}}{\sqrt{N}} \sqrt{1 - \frac{1}{1 + \frac{p_x \eta}{\xi} \frac{1}{2r} \left(\hat{T}_c \left(\frac{p_x \eta}{r \xi} \right) - N \right)} \right|^2}{1 + \frac{\frac{p_x}{\xi} \left(1 + \frac{\eta e^{-\eta} - \xi e^{-\xi}}{e^{-\eta} - e^{-\xi}} \right)}{1 + \frac{p_x \eta}{\xi} \frac{1}{2r} \left(\hat{T}_c \left(\frac{p_x \eta}{r \xi} \right) - N \right)}} \right) & r < N \\ \log \left(1 + \frac{\frac{N p_x p_z}{\xi} \left| E[\tilde{\mathbf{h}}_{0,0}] + \frac{\mathbf{u}}{\sqrt{N}} \sqrt{1 - \frac{1}{1 + \frac{p_x \eta}{\xi} \frac{1}{2L} \left(\hat{T}_c \left(\frac{p_x \eta}{\xi} \right) - N \right)} \right|^2}{1 + \frac{\frac{p_x}{\xi} \left(1 + \frac{\eta e^{-\eta} - \xi e^{-\xi}}{e^{-\eta} - e^{-\xi}} \right)}{1 + \frac{p_x \eta}{\xi} \frac{1}{2L} \left(\hat{T}_c \left(\frac{p_x \eta}{\xi} \right) - N \right)}} \right) & r = N \end{cases}, \quad (\text{D.8})$$

and $\mathbf{u} \sim \mathcal{CN}(0, 1)$ is a proper complex Gaussian random variable.

The proof of this bound follows the same lines as the proof of Theorem 3 with one major shortcut. We use the fact that the estimation error variance is a decreasing function of each of the transmitted symbols' power. We thus upper bound it by the estimation error variance that is achieved when all previous symbol used the minimal allowed transmission power. Taking into account the transmitted signal structure, (78), the bound results by replacing the term ζ by $\frac{c^2 \eta}{p_z}$ in (108) and (113).

As this bound does not use the convexity argument, (A.11) - (A.18), it does not require the zero mean assumption and holds for any expectation of the channel.

REFERENCES

- [1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [2] W. Kozek and A. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1579–1589, 1998.
- [3] G. Durisi, U. Schuster, H. Bölcskei, and S. Shamai, "Noncoherent capacity of underspread fading channels," *IEEE Trans. on Information Theory*, vol. 56, no. 1, pp. 367–395, 2010.
- [4] F. Hlawatsch and G. Matz, *Wireless Communications Over Rapidly Time-Varying Channels*. Academic Press, 2011.
- [5] M. Medard and R. G. Gallager, "Bandwidth Scaling for Fading Multipath Channels," *IEEE Trans. on Information Theory*, vol. 48, no. 4, pp. 840–852, Apr. 2002.
- [6] V. Sethuraman and B. Hajek, "Capacity Per Unit Energy of Fading Channels With a Peak Constraint," *IEEE Trans. on Information Theory*, vol. 51, no. 9, pp. 3102–3120, 2005.
- [7] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidot, "Low-SNR capacity of noncoherent fading channels," *IEEE Trans. on Information Theory*, vol. 55, no. 4, pp. 1555–1574, 2009.
- [8] I. E. Telatar and D. N. C. Tse, "Capacity and Mutual Information of Wideband Multipath Fading Channels," *IEEE Trans. on Information Theory*, vol. 46, no. 4, pp. 1384–1400, July 2000.
- [9] D. Porrat, D. Tse, and S. Nacu, "Channel Uncertainty in Ultra-Wideband Communication Systems," *IEEE Trans. on Information Theory*, vol. 53, no. 1, pp. 194–208, 2007.
- [10] J. Doob, *Stochastic processes*. Wiley, New York, 1953.
- [11] Y. Liang and V. V. Veeravalli, "Capacity of Noncoherent Time-selective Rayleigh-fading Channels," *IEEE Trans. on Information Theory*, vol. IT-50, no. 12, pp. 3095–3110, Dec. 2004.
- [12] A. Lapidot and S. M. Moser, "Capacity Bounds Via Duality With Applications to Multiple-Antenna Systems on Flat-Fading Channels," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2426–2443, Oct. 2003.
- [13] A. Lapidot, "On the Asymptotic Capacity of Stationary Gaussian Fading Channels," *IEEE Trans. on Information Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
- [14] T. Koch and A. Lapidot, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," in *Proceedings of the International Symposium on Information Theory, ISIT*, 2005, pp. 661–665.
- [15] J. G. Proakis, *Digital Communications*. McGraw-Hill, 1995.
- [16] H. Bölcskei, D. Gesbert, and A. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. on Communications*, vol. 50, no. 2, pp. 225–234, 2002.
- [17] D. Schafhuber and G. Matz, "MMSE and adaptive prediction of time-varying channels for OFDM systems," *IEEE Trans. on Wireless Communications*, vol. 4, no. 2, pp. 593–602, 2005.
- [18] M. Medard, "The Effect upon Channel Capacity in Wireless Communications of Imperfect Knowledge of the Channel," *IEEE Trans. on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [19] F. Neeser and J. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.
- [20] E. Baccarelli and A. Fasano, "Some simple bounds on the symmetric capacity and outage probability for QAM wireless channels with Rice and Nakagami fadings," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 361–368, 2000.
- [21] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. on Information Theory*, vol. 45, no. 1, pp. 139–157, 1999.

- [22] I. Bergel and S. Benedetto, “The effective coherence time of common channel models,” in *IEEE 11th Workshop on Signal Processing Advances in Wireless Communications. SPAWC*, 2010.
- [23] R. Clarke, “A statistical theory of mobile-radio reception,” *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, 1968.
- [24] A. Lapidoth and S. Shamai, “Fading Channels: How Perfect Need “Perfect Side Information” Be?” *IEEE Trans. on Information Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [25] X. Jin, T. Li, O. Collins, and T. Fuja, “The universality of LDPC codes on correlated fading channels with decision feedback based receiver,” *IEEE Global Telecommunications Conference, GLOBECOM 2005*, vol. 3, 2005.
- [26] T. Li and O. Collins, “A successive decoding strategy for channels with memory,” *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 628–646, 2007.
- [27] G. Durisi, V. Morgenshtern, and H. Bölcskei, “On the Sensitivity of Noncoherent Capacity to the Channel Model,” in *Proceedings of the International Symposium on Information Theory, ISIT*, 2009, pp. 2174–2178.
- [28] L. Zheng, D. Tse, and M. Medard, “Channel Coherence in the Low-SNR Regime,” *IEEE Trans. on Information Theory*, 2007.
- [29] J. Chen and V. Veeravalli, “Capacity Results for Block-Stationary Gaussian Fading Channels With a Peak Power Constraint,” *IEEE Trans. on Information Theory*, vol. 53, no. 12, pp. 4498–4520, 2007.
- [30] I. E. Telatar, “Capacity of Multi-Antenna Gaussian Channels,” *European Trans. on Telecommunications*, vol. 10, pp. 585–595, Nov. 1999.
- [31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.
- [32] S. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice Hall PTR, 1993.
- [33] M. Abramowitz and I. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Dover Publications, 1965.

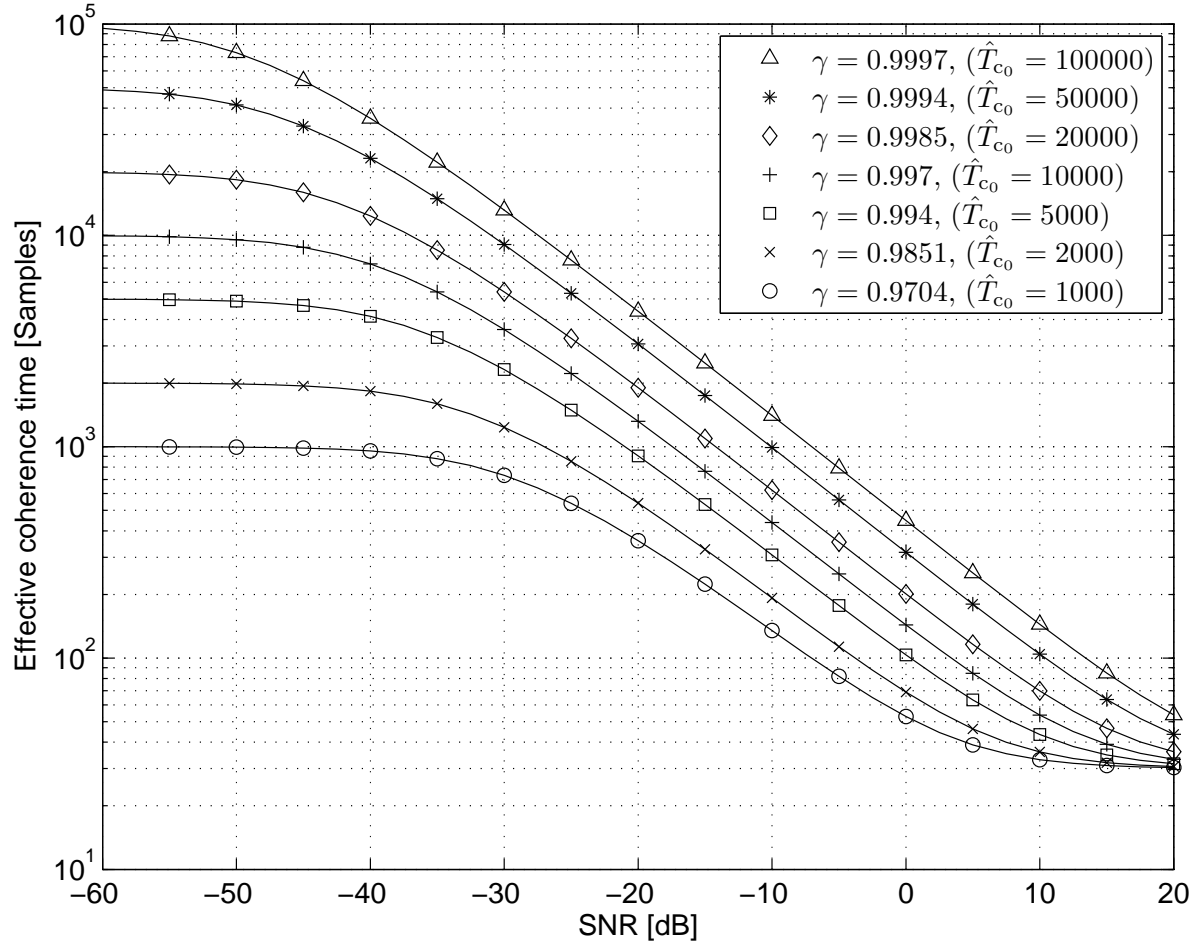


Fig. 1. Effective coherence time of the AR1 channel model for various values of the channel forgetting factor.

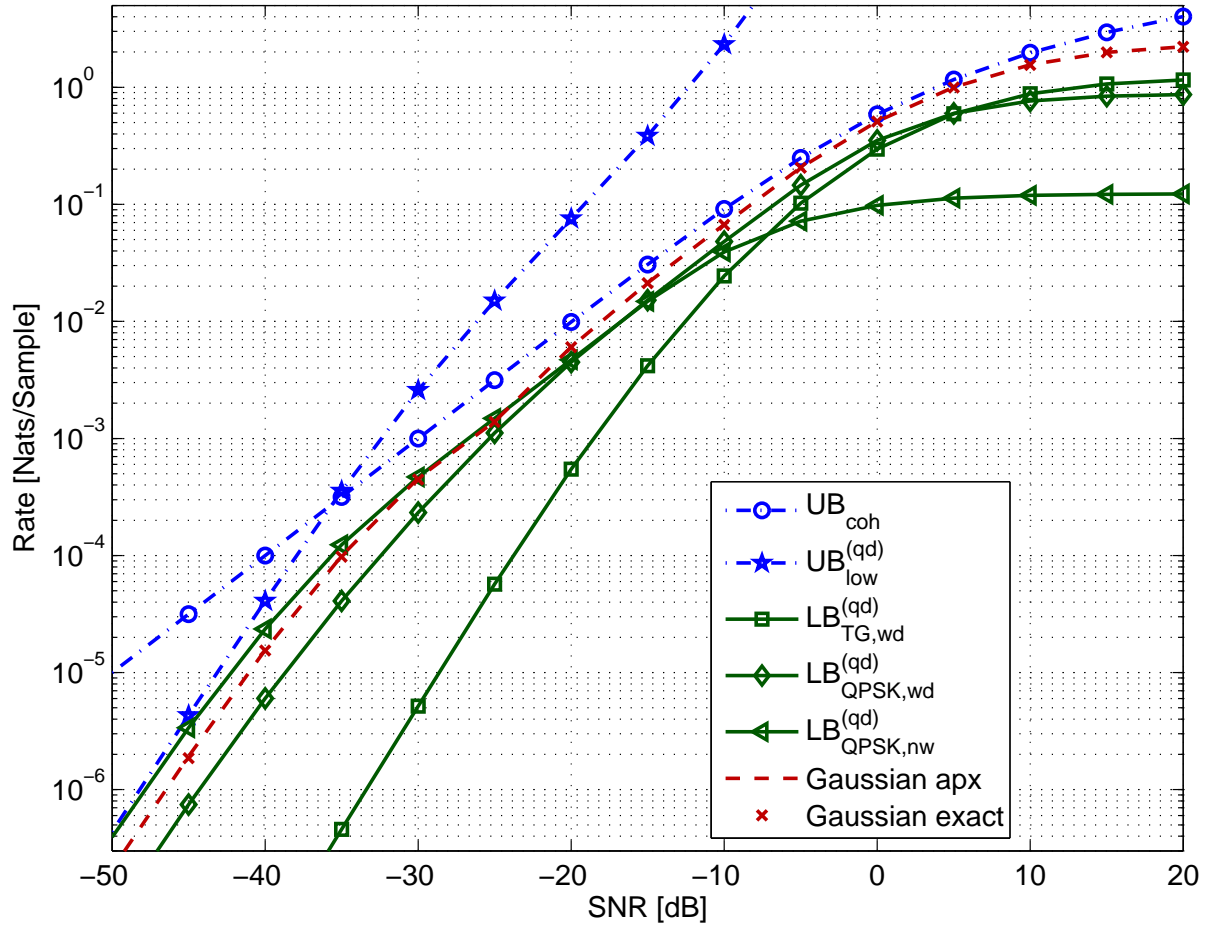


Fig. 2. Upper and lower capacity bounds vs. SNR for AR1 fading channel ($\hat{T}_{c_0} = 900$) and quadratic power constraint ($\alpha = 10$).

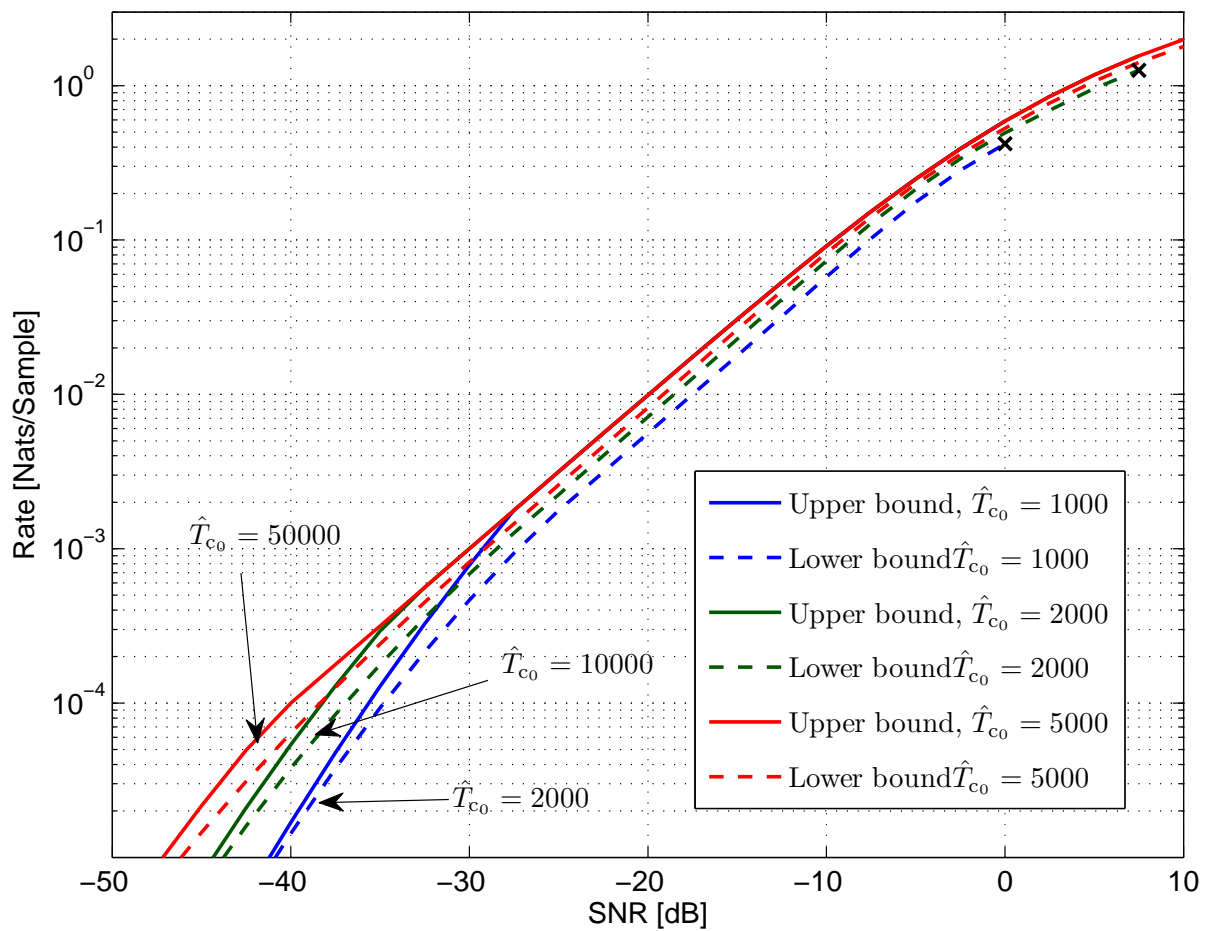


Fig. 3. Upper and lower capacity bounds vs. SNR for AR1 fading channel ($\hat{T}_{c_0} = 2000, 10000, 50000$) and quadratic power constraint ($\alpha = 2$).